Winter 1993

The association properties of retroviral integration protein

Jacqueline Lei Coleman
University of New Hampshire, Durham

Order Number 9420562

The association properties of retroviral integration protein

Coleman, Jacqueline Lei, Ph.D.

University of New Hampshire, 1993

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

THE ASSOCIATION PROPERTIES OF RETROVIRAL INTEGRATION PROTEIN

BY

JACQUELINE L. COLEMAN
B.A., University of North Carolina, 1973
M.S., University of Maine, 1981

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

in

Biochemistry

December, 1993

This dissertation has been examined and approved.

Dissertation Director, Thomas M. Laue
Associate Professor of Biochemistry

Rick H. Cote
Assistant Professor of Biochemistry

Andrew P. Laudano
Associate Professor of Biochemistry

Howard R. Mayne
Associate Professor of Chemistry

Charles K. Zercher
Assistant Professor of Chemistry

October 25, 1993

ACKNOWLEDGEMENTS

I decided to pursue my Ph.D. in Biochemistry not long before my 39th birthday. The impact of my decision to return to school was probably greater on my family than it was on me. Yet their support of and faith in me never faltered. Never once did I hear the words "midlife crisis" uttered in my home. Whenever I found myself heading upstream, never once did they fail to provide me with a paddle. I do not possess the vocabulary which enables me to tell them what their unconditional love and support mean to me. All I can do is acknowledge Mom, Dad and Helene.

I must also thank Tom Laue and Terry Ridgeway not only for the friendship and support that they show all members of their laboratory, but for the extra support which was given to me when needed. Graduate students, like children, leave their laboratory family and forget to call or write. Yet Terry and Tom will forever be in my thoughts.

I would like to thank Gene Putz, my friend and mentor in the clinical laboratory for many years. His unyielding faith in me helped keep me focused on my goals.

I have enjoyed the friendship of many graduate students at UNH, but I will always carry special memories of Craig Almeida and Tom Moody. Their support as well as the support of many of my friends will never be forgotten.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

THE ASSOCIATION PROPERTIES OF RETROVIRAL INTEGRATION PROTEIN

by

Jacqueline L. Coleman
University of New Hampshire, December, 1993

Retroviral Integration Protein (IN) has been shown to be both necessary and sufficient for the integration of reverse transcribed retroviral DNA into the host cell DNA. It has been demonstrated that self-association is required for proper function, yet little is known about the thermodynamic properties of this association. Analytical ultracentrifugation was used to determine the stoichiometries and free energies (ΔG) of IN self association under various physical conditions. The best fit model of association in all glycerol-free conditions studied was a reversible monomer-dimer-tetramer. The addition of glycerol at concentrations of 8% (vol/vol) or greater significantly enhances the monomer-dimer formation so that the best fit model is dimer-tetramer.

The data support a model in which distinct binding sites on the protein are responsible for monomer-dimer and dimer-tetramer association. The strength of the association is influenced by hydrogen bonding as well as charge-charge, dipole and hydrophobic interactions. The understanding of the thermodynamics of IN association may aid in targeting this protein for therapeutic intervention.

INTRODUCTION

Retroviruses comprise a relatively homogeneous class of organisms distinguished by the fact that they are dependent upon a DNA integration step for replication. Retroviral integration protein (IN) has been shown to be both necessary and sufficient for this integration step, yet the precise mechanism of this protein is unknown (for review, see 1-5). It is the study of a portion of this mechanism that is the focus of this research.

This section will review the life cycle of the retrovirus and then focus on the integration step, including a discussion of the mechanism and energy requirements of integration. The methods used for this study will be presented, followed by the results obtained from the study of the association properties of IN under different physical conditions. These results will then be discussed and an integration model presented that is supported by the results.

The general description of a retrovirus particle (virion) is a 90-100 nm sphere consisting of a nucleoprotein core which is wrapped in an envelope obtained by budding from the host cell surface. The nucleoprotein core consists of a 7-9 kilobase RNA genome and several viral encoded proteins. These proteins include structural components, a transmembrane complex and enzymes for proteolysis (PR), reverse transcription (RT) and integration (IN) (6). The core may

also contain a variety of other components derived from the host, some of which may play a role in the life cycle of the virus and some of which have a purpose that has yet to be described.

Retroviruses are unique among animal viruses in that they require integration of their DNA into the host cell's for efficient replication. This feature of their life cycle is even more unique than the reverse transcription for which they are named, for reverse transcription is a characteristic shared with the hepadnaviruses and caulimoviruses. Retroviruses have two identical RNA subunits contained in their nucleoprotein core. This feature also distinguishes the retroviruses from all other animal viruses, which are haploid (5).

There are many reasons for interest in retroviruses. These biochemically unique viruses have been found in all vertebrate animals in which they have been sought (5). Studies of retroviruses have shown that they provide clues to early evolutionary events, as well as to genetic feats such as insertional mutagenesis. A great deal of interest has focused on this class of viruses, most recently due to the fact that Human Immunodeficiency Virus (HIV), the causative agent of AIDS, is a retrovirus. Retroviruses include a class of carcinogenic agents known as oncogenes. One of the members of this oncogenic class is Rous Sarcoma Virus (RSV) (for review,

see 7-8). It is RSV IN that is used as the integration protein studied in this project.

A review of the literature shows that most of the work on the integration of the retrovirus into the host cell has been done using integration proteins from three different retroviruses: 1. Rous Sarcoma Virus (RSV), which is a representative of the avian retrovirus class. The original work which described the protein later to be known as IN was performed on this class of retrovirus (9), 2. Murine Leukemia Virus (MuLV), 3. Human Immunodeficiency Virus (HIV), of which HIV-1 IN is the protein most commonly used. The IN used in these experiments has come from both virion and bacterial (plasmid) preparations (10). While there are some differences in the integration protein of these viruses, the general model of integration is common to all retroviruses studied to date (2, 10-13).

Historically, retroviruses have been so named because of their dependence upon the enzyme reverse transcriptase to transcribe a double-stranded DNA intermediate from the single-stranded RNA carried in their genome. As mentioned above, this reverse transcription event is shared with other viral classes. The single lifecycle event that distinguishes retroviruses from other viruses is their dependence upon integration for replication. The complete life cycle of the retrovirus is depicted in Figure 1 (5). The steps in this life cycle are as follows: 1. an enveloped virus attaches to

FIGURE 1

RETROVIRAL LIFE CYCLE

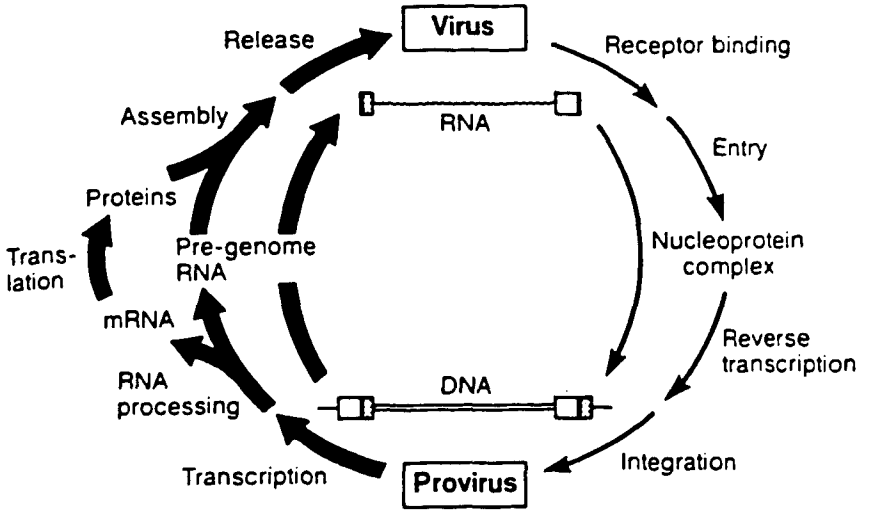


Figure 1a. An outline of the steps in the retroviral life cycle (5).

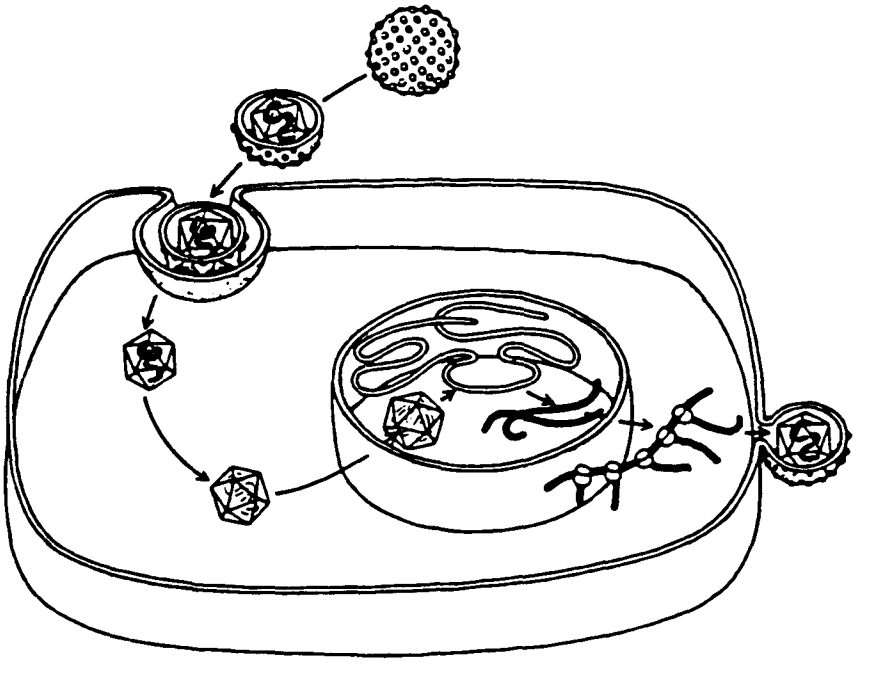


Figure 1b. A pictorial overview of the steps outlined above (5).

a transmembrane receptor of a susceptible cell, 2. the virus sheds its outer protein coat (envelope) and enters the cell as a nucleoprotein complex, probably via receptor-mediated endocytosis, 3. the single-stranded viral RNA is transcribed to double-stranded DNA by the enzyme reverse transcriptase, 4. migration of the DNA-protein complex into the host nucleus, 5. processing and joining of the host and viral DNA by retroviral integration protein, forming the provirus, (this step may begin prior to migration into the host nucleus), 6. replication and transmission of provirus to daughter cells as a part of the host DNA replication, 7. transcription of DNA to RNA to be used for mRNA and progeny genome, 8. processing and translation of viral RNA, 9. modification and assembly of protein products, 10. assembly of the viral nucleoprotein core, and 11. budding through plasma membrane. This step occurs in a glycoprotein-rich region so that the nucleoprotein complex acquires its outer coat (envelope) to form a virus particle.

It is important to note three things about this life cycle as described. First, as with all life cycles, the beginning and end are arbitrarily assigned. Second, there is a division between those events that are dependent upon virus-encoded proteins which enter the host cell with the viral RNA, which include reverse transcriptase and IN, and those events dependent upon the host cellular machinery to continue the life cycle. Third, the processing and joining reactions of

the integration step are unique in that the integration protein may function in both the cytoplasm and nucleus of the host cell.

The integration step in the life cycle is the focus of this research. It is this step that enables the virus to begin the multiplication activities vital to its survival through progeny. Integration makes the virus a stable part of the dividing host cell, and once the virus is integrated it will be transmitted every time the host cell divides. It is also this integration step that gives the retrovirus several unique properties, including insertional mutagenesis, oncogenic transformation, latency, persistence of infection, and usage as genetic engineering vectors (1).

It is now known that the integration reaction requires two factors contained in the viral nucleoprotein complex (1,14). The first of these factors is a product of reverse transcription. During this process, terminal redundant double stranded DNA copies of the viral genome are produced which are called long terminal repeats (LTRs) (for review, see 15,16). LTRs are cis-acting sequences required for integration (17-21). The second required factor is a trans-acting protein that is a product of the viral polymerase (*pol*) gene. This protein is IN (9,11,22-26). IN is encoded in the 3′ portion of the *pol* gene as a larger precursor and then proteolytically processed at both the N- and C- terminals to a smaller subunit (27). HIV and RSV IN molecules are 32,000 daltons, and the

MuLV IN is a 46,000 dalton protein (2).

The first step in the integration reaction is the cleavage of the 3′ ends of the double-stranded viral DNA by IN. This cleaved DNA is detected in the cytoplasm of host cells within a few hours of viral invasion (21,28). The ends of linear DNA produced during reverse transcription are blunt, ending in the sequence 5′-AATG.....CATT-3′ (29). It has been demonstrated that the double stranded viral DNA is trimmed on one strand at each end and that a two nucleotide product can be detected (10,21,30,31). The factor which determines the point of cleavage of viral DNA is not the number of bases removed, but the location of the highly conserved CA dinucleotides located at the 3′ end of each of the strands of the viral DNA (3,17,18,21,28,32). Data show that unintegrated linear viral DNA from HIV-1, MLV and RSV all likely contain two nucleotides beyond this conserved region (33). Mutation studies have indicated that, although cleavage always takes place at the conserved CA dinucleotide, up to 15 bases internal to the CA region are required for proper cleavage (15,18,19,34,35). Single point mutations of bases in this region, however, do not appear to have a significant effect on cleavage (18,34,35). Since intermediates lacking one 3′ nucleotide are not found, it appears that a single cleavage event removes both terminal nucleotides as a unit (21). These base pairs are nicked through the endonuclease activity of IN in the cytoplasm, and it has been proven that IN alone is

necessary and sufficient to perform this cleavage, now known as the processing reaction (19,34,35). It also has been indicated that the reaction may be species-specific; recombinant HIV-1 IN that was produced in *E. Coli* effectively cleaved HIV-1 substrates, but not substrates of other species (32).

The next step in the integration reaction is the transport of the processed viral DNA to the host cell nucleus. While little is known about this transport mechanism, it is thought that IN plays an active role. It has been suggested that IN may carry a signal for nuclear localization (36,37).

The final integration step is a two-step reaction in which the host cell DNA is cleaved and the viral DNA inserted. It has been proven that IN alone is necessary and sufficient for this joining process (10,22). In this step, the processed 3′ ends of the viral DNA are inserted into host DNA by a mechanism to be discussed later. The integrated viral DNA, now known as a provirus, is joined at the cleaved 3′CA dinucleotide in its LTR site to the 5′ end of the host cell DNA. The 5′ end of the viral DNA remains unjoined. The mechanism of excision of the unjoined viral nucleotides as well as the repair of the gap created by host cell cleavage remains unknown. The gapped intermediate which is created during the joining reaction is uniform in structure. The LTRs flank an identifiable coding region, generally 2 to 8 bases and invariant in length with regard to species (5,31). It is

thought that the excision and gap repair steps are accomplished by host cell enzymes (5,38,39). The complete integration reaction is illustrated in Figure 2 (4).

The location of proviral integration in the host cell DNA has been a recent subject of interest. The retroviral provirus is found at a variety of sites in the host genome. However, no sequence specificity has been found (18,22). It has been shown on the other hand that there appear to be "hot spots" where the provirus appears more often than expected in a random reaction (40). Studies also provide evidence that preferential target sites are associated with transcriptionally active sites (41,42) and with the major groove on chromatin assembly (13). This finding is surprising because it is thought that the presence of histones inhibits DNA activity (for review, see 43,44). It has been proposed that proteins bound to target DNA may be recognized by IN, directing the IN-viral DNA complex to specific sites (1). Since DNA-bound proteins tend to be very positively charged, it would require a significant binding energy between IN and target DNA to overcome the charge-charge repulsion between these two molecules.

The insertion of the provirus is permanent and vital to the efficient expression of progeny viruses. While evidence has been provided that unintegrated retrovirus DNA can be utilized as a template for viral protein production, the unintegrated virus is not infectious (20,45). Once the

FIGURE 2

MODEL OF RETROVIRAL INTEGRATION

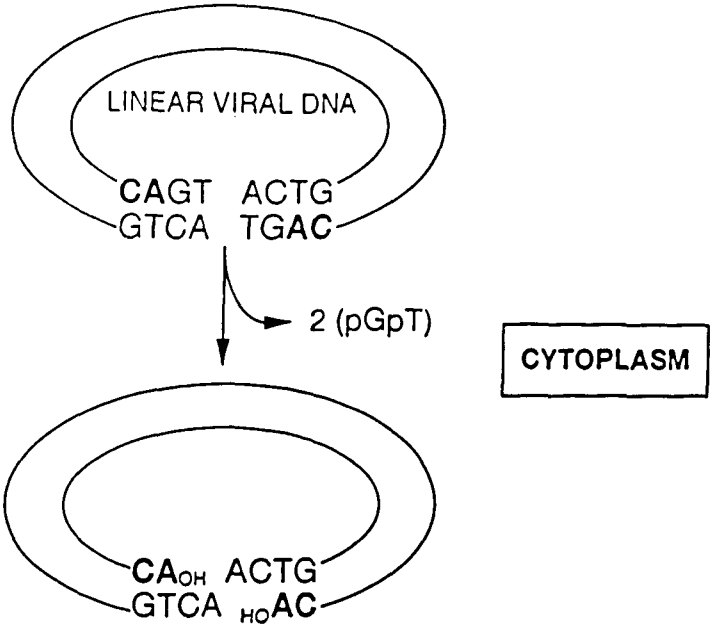


Figure 2a. Processing Activity. This occurs in the cytoplasm of the host cell (4).

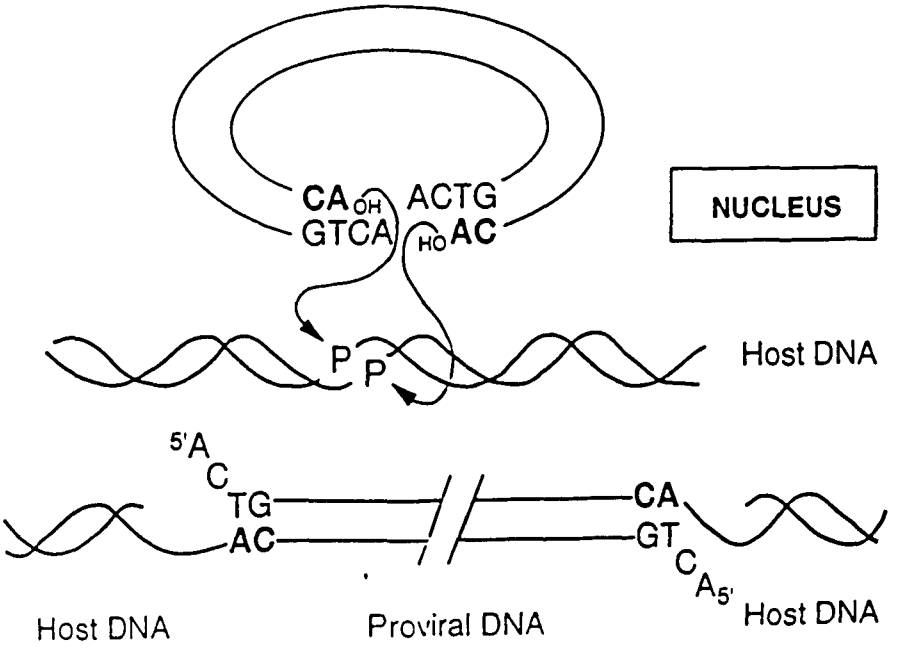


Figure 2b. Joining Reaction. This takes place in the nucleus of the host cell (4).

provirus of one invading retrovirus is integrated into the host, there may be thousands of virus particles produced per cell per day as the single provirus is copied by transcription of host cell DNA. Since retroviruses rarely kill the cells they infect, this amplification potential is enormous (5).

Recently, another enzymatic reaction of IN has been described (38). It has been shown that the cleavage-ligation joining step is reversible, a reaction referred to as disintegration. It is not known what significance this process may have *in vivo*.

While the exact mechanism by which the processing and joining reaction takes place is not known, it is known that the reactions attributed to IN require no exogenous energy source (1-5,10). IN does require a divalent cation for proper function. When the cation is Mg^{2+}, IN is very specific in its cleavage of the viral DNA. When Mn^{2+} is the cation, IN activity is greatly enhanced but an additional major cleavage site at the -3 position from the 3′ end becomes apparent (19). It is not known if this extra product is of any significance *in vivo*.

Viral DNA is present in cells in three forms: 1. linear DNA, 2. circular DNA with one LTR, and 3. circular DNA with two LTRs. While an initial study indicated that circular DNA was the integrated form (46) it is now widely accepted that the precursor to the provirus is the linear form (10,15,28,47-49). The reason for more than one form of viral DNA has not

yet been explained.

Enzymatic Mechanism of Integration

The initial processing reaction of the viral DNA by IN at the conserved CA dinucleotide occurs through the endonuclease activity of IN. IN nicks the ends of the virus while still in the cytoplasm. The mechanism involves site-specific hydrolysis of a phosphodiester bond. No evidence has been found for any covalent linkage in this step, suggesting a direct nucleophilic attack of water at this phosphodiester bond immediately 3′ adjacent of the conserved CA(TT)3′ (10,50). This mechanism is supported by the fact that pTpT dinucleotide (or equivalent) is the major product (51). Further evidence for this mechanism is provided by the fact that a "precleaved" substrate with a 3′OH end is efficiently joined to target DNA (34). While it has been demonstrated that a variety of alcohols, including glycerol and alcoholic amino acids, can function as the nucleophile *in vitro*, no evidence has been found that the IN protein is covalently attached to the pTpT (or equivalent) dinucleotide. Therefore, an IN-directed single nucleophilic attack of water at the specific phosphodiester bond appears to be the most likely mechanism (51). This step in the reaction is shown in Figure 3a (5).

It has been proven that the second half of the IN reaction, that of processing the host DNA and joining the viral DNA to the host DNA, is independent of the processing of

the viral DNA described above (5,10,15,32,35,52). Linear viral DNA with preprocessed ends still can be integrated into host DNA without an outside energy source. It is therefore not likely that the energy source of the second half of the reaction is provided by a high-energy intermediate generated in the first half of the reaction.

Two possible mechanisms have been proposed for the second half of the IN reaction. Both have the energy for strand transfer derived from the phosphodiester bonds broken in the host cell (target) DNA in which the ultimate nucleophile is the 3′ hyroxyl oxygen at the preprocessed end of the viral DNA. This oxygen is joined to the 5′ phosphate of the cleaved target DNA (51,15). In this strand transfer reaction, the 3′ end of the cleaved target DNA strand is unjoined. It is proposed that repair of the gap created by this unjoined 3′ end is performed by host enzymes and is independent of IN activity.

One mechanism proposes the formation of a high energy covalent bond between IN and the 5′ phosphate ends of target DNA produced at the sites of cleavage. This bond may involve a conserved serine or threonine. This model proposes that the conserved hydroxyl oxygen on the alcoholic amino acid would be attacked by the nucleophilic oxygen on the 3′ ends of the cleaved viral DNA (5,51-53). This mechanism is illustrated in Figure 3b (5).

The second, more likely, model involves a concerted one-

step mechanism in which the 3'OH end of the processed viral DNA acts as the nucleophile to attack a phosphodiester bond at the site of the target DNA. A stereochemical analysis of this reaction has demonstrated that this step occurs with an inversion of configuration. The two-step mechanism in which a high-energy intermediate is utilized would lead to a retention of configuration (52). The one-step model is illustrated in Figure 3c (5).

Structure of IN

Three separate regions have been characterized on IN: 1. a zinc-finger containing N-terminal region, 2. a conserved catalytic core region, and 3. a non-conserved C-terminal region. Computer-assisted analysis of more than 80 retroviruses, retrotransposons and bacterial insertion sequence transposases has shown eleven highly conserved amino acids (54). Figure 4a is the complete 286 amino acid sequence of RSV IN with the conserved amino acids indicated (27). Figure 4b is a picture of the location of the conserved amino acids on a linear model of IN. The CO4 mutant described in the results section is depicted also (51,72). The effect of point mutations of these conserved amino acids as well as the effect of deletion mutants on IN activity have been the subjects of several recent studies.

No mutations studied to date have resulted in a significant difference in the effect on cleavage and joining. A mutation which results in poor cleavage also shows poor

FIGURE 3

MECHANISM OF RETROVIRAL INTEGRATION

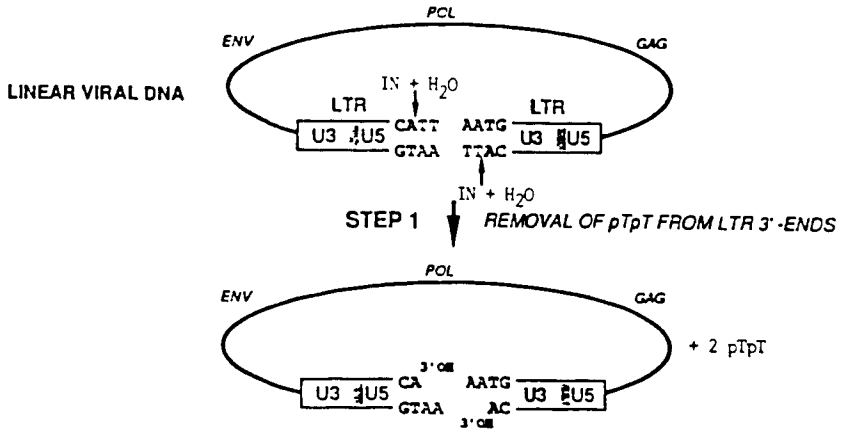


Figure 3a. Processing Activity. This occurs in the cytoplasm of the host cell.(10)

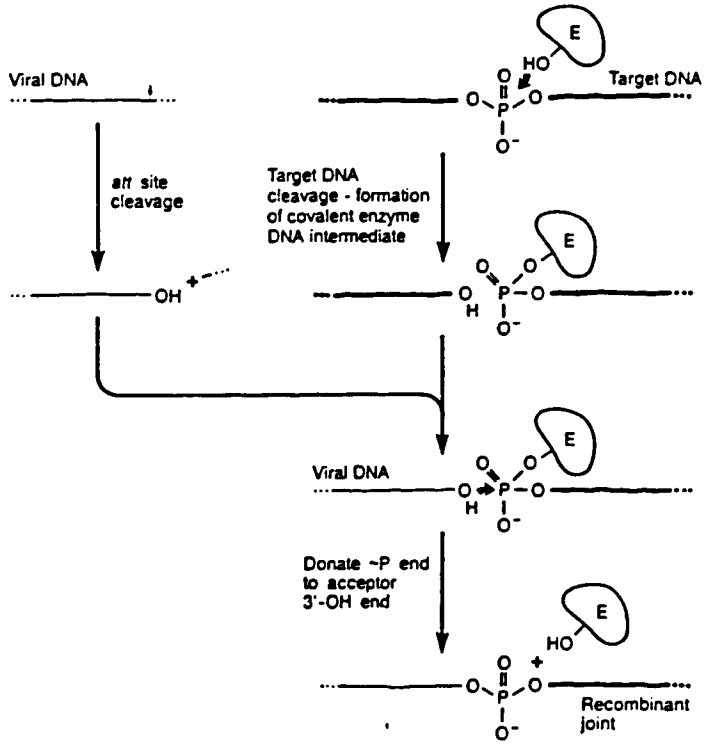


Figure 3b. Joining Reaction which uses a covalent intermediate. Only the events on one strand are shown(5).

FIGURE 3

MECHANISM OF RETROVIRAL INTEGRATION

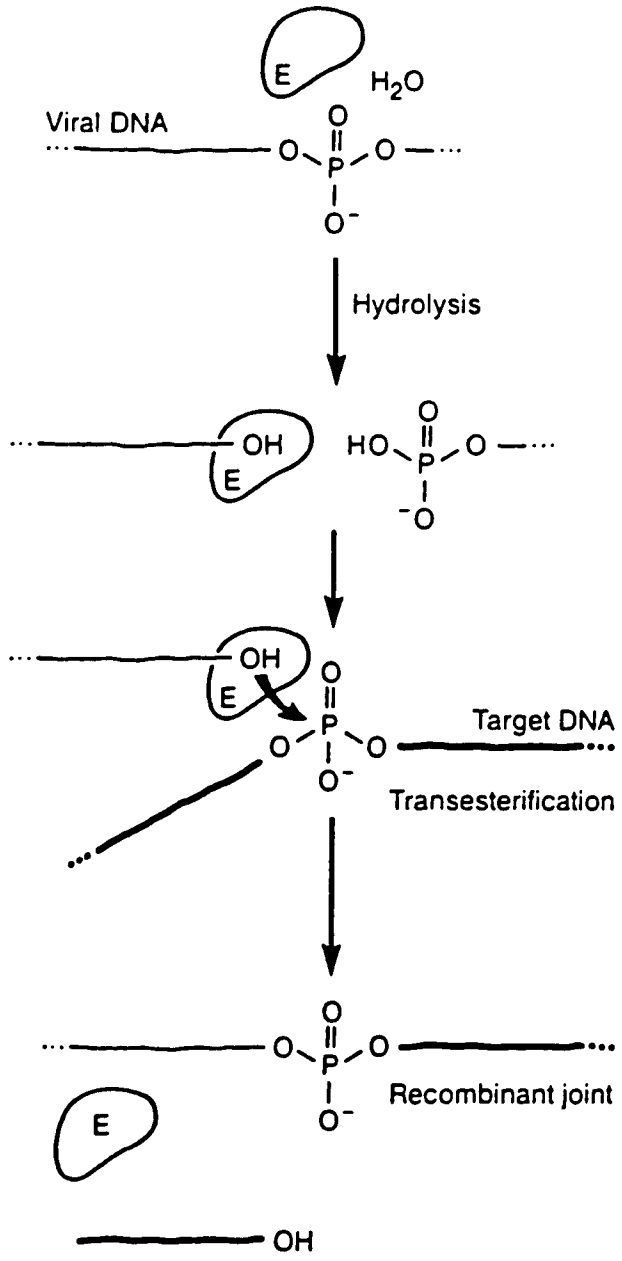


Figure 3c. One-step Joining Reaction Mechanism. Only the events on one strand are shown(5).

FIGURE 4

P L R E A K D L H T A L H I G P R A L S K A C N I S M Q Q A R

(32)E V V Q T C P H C N S A P A L E A G V N P R G L G P L Q I

W Q T D F T L E P R M A P R S W L A V T V D T A S S A I V V T

(92)Q H G R V T S V A V Q H H W A T A I A V L G R P K A I K T

D N G S C F T S K S T R E W L A R W G I A H T T G I P G N S Q

(152)G Q A M V E R A N R L L K D R I R V L A E G D G F M K R

I P T S K Q G E L L A K A M Y A L N H F E R G E N T K T P I Q

(211)K H W R P T V L T E G P P V K I R I E T G E W E K G W N

V L V W G R G Y A A V K N R D T D K V I W V P S R K V K P D I

(270)T Q K D E V T K K D E A S P L F A

Figure 4a. The amino acid sequence of RSV IN. The eleven conserved amino acids are underlined (27).

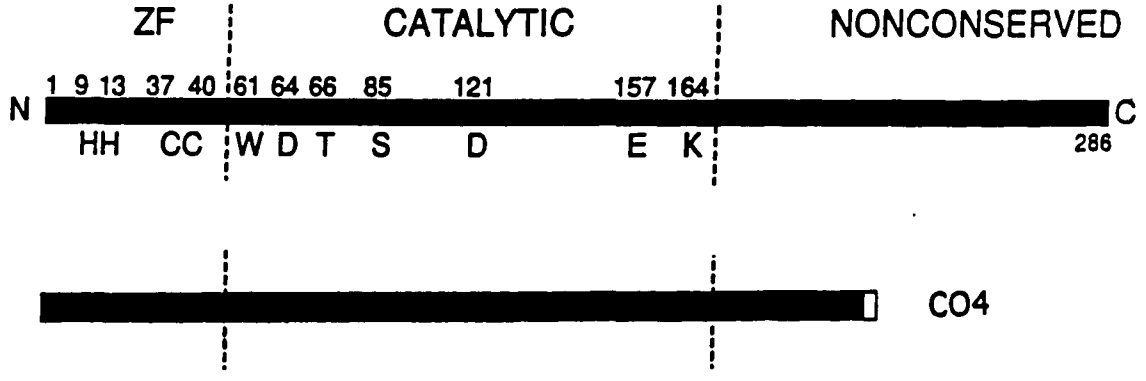


Figure 4b. A linear model of the RSV IN molecule. The conserved amino acids are indicated in the wild-type model at the top (51,72). The CO4 deletion mutant is shown at the bottom.
ZF = Zinc Finger

joining activity, and conversely any mutation which shows effective cleavage also shows effective joining activity (15). This suggests that the IN mechanism may be similar for these two steps.

The N-terminal region, also referred to as the HHCC or Zinc-finger region, has four conserved amino acids. In RSV IN this region is defined by residues 9 through 40 and the conserved amino acids are His-9, His-13, Cys-37 and Cys-40. It has been proposed that the conserved HHCC region comprises a zinc-finger region necessary for DNA binding (27,55,56). It has been shown that IN can bind zinc and that the HHCC region is a requirement for the zinc binding (55). The region varies from the prototype zinc finger, however, in that the HHCC region of IN is about twice the size and much longer than zinc-fingers known to bind DNA. The order of the residues is also different from that found in other zinc fingers, although a conformational "twist" could result in an apparent change of this sequence (27). It has been shown, however, that the conserved HHCC region is not required for DNA binding activity. Instead, this region contributes to specific recognition of the viral LTR and therefore it is proposed that these parts may interact (27,55).

The second described region, the catalytic or DDE region, is the most highly conserved among the retroviruses. Three residues in this region of RSV IN have been the focus of recent studies: Asp-64, Asp-121 and Glu-157. Data show that

the HHCC and DDE regions probably lie on separate independently folded domains (55). It has been proposed that the DDE region is the catalytic region for all IN enzymatic activities identified to date (55,57). All substitutions of the invariant D and E residues reduce processing to a few percent or less of wild type. The ability of IN to bind and cleave DNA is virtually destroyed (27,54). It is proposed that these residues may participate in the coordination of the divalent cation required for the catalytic activities of IN. This coordination may be required in order to position both the viral and host DNA for nucleophilic attack during the cleavage and joining reactions (54). Another study demonstrates that the point mutations of two nonconserved negatively charged amino acids in this region results in complete or nearly complete loss of 3′ processing and joining activity (56). Point mutations of the conserved Ser-85 in this region also result in severe reduction of both processing and joining activities of RSV IN (51). This same study shows that a point mutation on the conserved Thr-66 does not significantly reduce RSV IN enzymatic activity.

The carboxy-terminal portion of IN is a nonconserved region. Deletion mutations have shown, however, that this region is important in DNA binding (27,57). The role of this region may be to aid in the formation of IN oligomers which are necessary for DNA binding and enzymatic activity (55). One interesting study in particular was performed in this

region of HIV IN. While mutating Arg-199 to a cysteine, a frameshift mutation occurred which lacked the carboxy terminal 42 residues of wild type but acquired 19 unrelated amino acids. The 3' processing activity of this mutation was at most mildly impaired, but the strand transfer activity was markedly reduced. Disintegration activity, however, was the same as wild type. Thus, strand transfer and disintegration, two IN activities which are thought to be just the enzymatic reversal of each other, were differentially affected (56). A possible explanation of this observation could be that one activity requires oligomerization and the other does not.

Two separate studies have been performed recently on HIV mutants that show when one of the three major domains is mutated, it can be complemented with a second mutant deficient in another domain but with the first domain intact. Thus, when the activity of the two mutants mixed together is measured it is the same as wild type (39,58). This complementation works best when the wild type N- and C-termini are present on the same molecule. This data support a model of multimerization and show that not all sites are necessary for proper function of the multimer.

It has been proposed in many studies that IN probably functions as a multimer (2,11,12,27,51,57,59). An interesting observation that must be included in any proposed model is that replication-defective viral mutants with deletions at one edge of the DNA substrate are not processed at the wild-type

end (21,59). This suggests that two functional viral termini are required for efficient integration, and the processing of these termini may be related to the IN conformation required for multimerization. The previously discussed complementation studies suggest that the sites for possible DNA binding (the DDE domain) and protein oligomerization (the carboxy-terminus) are separate. Thus one could envision a model where IN multimerization could occur while allowing a site to be conformationally available for DNA binding.

The purpose of this research is to establish the stoichiometry and free energy of IN self-association. The IN molecule was studied using the analytical ultracentrifuge. This method is chosen because it allows the study of the thermodynamic and hydrodynamic properties of the protein and substrate in their native state. Data from the centrifuge were analyzed using NONLIN (60). The thermodynamic parameters then were determined under varied physical conditions. A set of molecules which have point mutations of some of the conserved amino acids then were examined and their association models compared to wild-type IN. The results of these findings then were used to propose a model of IN function.

MATERIALS AND METHODS

Materials

All RSV and HIV integration protein (IN) used in these experiments was provided by the laboratory of Dr. Anna Marie Skalka of the Fox Chase Cancer Center, Institute for Cancer Research, Philadelphia, PA. The protein was purified as defined in the method given in Appendix C. The protein was stored at -20 degrees C in a buffer of 50 mM TRIS (pH 7.5), 0.5 M NaCl, 2 mM DTT, 0.1 mM EDTA and 40% glycerol. The protein used was from pools labelled prep 15 and prep 19.

Sterile Centrex cellulose acetate columns for the centrifugal gel exchange were purchased from Schleicher and Schuell, Keene, N.H. Sephadex G-25 beads were purchased from Pharmacia. Anhydrous glycerol, 99.6%, was purchased from the Ultrapure Bioreagent line of J.T. Baker. All other components were ultrapure or reagent grade.

Buffer Exchange

Buffer exchanges for all experiments were performed by centrifugal gel chromatography (61). G-25 Sephadex beads used in the centrifuge columns were swollen either overnight at room temperature or for 2 hours in a boiling water bath (62). The columns were poured and equilibrated with at least 5 bed volumes of buffer prior to use. The columns were then centrifuged and the supernatant saved to be used as the

reference in the ultracentrifuge cells. The protein in its storage buffer was then applied to the column and centrifuged. The supernatant was used as the protein preparation in each of the experiments described in the results section. Protein dilutions were made with the same buffer preparation as that used for the centrifugal gel exchange. The complete procedure is described in Appendix B.

Sedimentation Equilibrium Analysis

Sedimentation equilibrium experiments were conducted on the Beckman Model E analytical ultracentrifuge at a temperature of 23.3 degrees C (63) with an on-line Rayleigh interferometer (64) using a pulsed laser diode light source at 670 nm. Rotor speeds for all experiments were 20,000, 24,000, 28,000 and 36,000. Either an AN-D (2 hole), AN-F (4 hole) or AN-G (6 hole) rotor was used in all experiments.

All protein solutions were spun in a high speed Beckman microfuge for 5 minutes to remove any particulate matter prior to cell loading. Four solution/solvent pairs were examined in each experiment using a 12 mm thick charcoal-filled epon or Kel-F centerpiece with sapphire windows. Unless otherwise indicated, all protein concentrations were 1.2 mg/ml in the first channel and diluted with buffer from the same centrifugal gel column for the remaining three channels. Protein concentrations were estimated using the extinction coefficient at 280 nm of 1.83 as determined by the tyrosine and tryptophan content of the protein (65).

Equilibrium was established by subtracting the fringe displacement from consecutive scans (made at least 15 minutes apart) of the most concentrated sample until there was no change from scan to scan. If the rotor contained cells with buffers of different ionic strengths, the cell with the highest concentration of salt was used for equilibrium determination. The average equilibrium time for solutions without glycerol was 1½ hours. The data reported from the experiments which contain glycerol were from experiments in which the cells were centrifuged at each speed for 24 hours before being tested for equilibrium. Equilibrium was determined in these cells by subtracting the fringe displacement from consecutive scans made at least one hour apart of the most concentrated sample until there was no change from scan to scan.

The experiments to determine the concentration and mole fraction of the multimeric forms of IN as a function of total concentration (Figures 5 and 6) included data from dilute solutions of IN. The cell used for these experiments was a specially constructed 37-μl sample volume, 30-mm pathlength, Kel-F short column centerpiece. The initial loading concentration was 0.29 mg/ml of protein equilibrated by centrifugal gel chromatography in a buffer of 0.5 M NaCl, 20 mM TRIS (pH 8.4) and 2 mM β-mercaptoethanol (BME). Dilutions of the protein were made using buffer from the same centrifugal column. Equilibrium determination and data

collection were performed as described.

The HIV IN experiments were performed using 0.4 mg/ml of protein equilibrated by centrifugal gel exchange in a buffer of 20 mM TRIS, 1 mM BME, 1 mM EDTA and either 1 M NaCl at a pH of 7.0 or 7.4 or 0.5 M NaCl at a pH of 7.4. The 12-mm short column centerpiece was used for this experiment. In addition, an experiment using the 30 mm centerpiece was also performed in a buffer of 0.5 M NaCl, 20 mM TRIS (8.4), 1 mM EDTA and 1 mM BME equilibrated by centrifugal gel chromatography. All other parameters of the experiment were as described above.

Buffer Density

The buffer density was determined in order to perform the calculations necessary for the analysis of the sedimentation data. The density of the individual components at 20 degrees C was obtained from published tables (66). This density was then subtracted from the density of water at 20 degrees C, which is .99823 g/ml. The relative contribution of each buffer constituent was determined, and the total solvent density calculated. This density was then multiplied by .999, which is the correction factor to 23.3 degrees C, the temperature of the centrifuge.

Partial Specific Volume

The partial specific volume ($\bar{V}$), which is a measure of the change in the volume (in ml) of the solution per gram of solute, was also an important parameter used in the

sedimentation calculations (67). The $\bar{v}$ measurements were performed using the VAX program VBAR (kindly provided by David Yphantis) which calculated $\bar{v}$ based on the protein's amino acid composition (67). The $\bar{v}$ calculated for IN using its amino acid sequence was 0.73_{76} ml/g.

Charge on IN

The charge on the protein was also calculated as a part of the VBAR program. This calculated charge is based on the pKa values of the individual amino acids. The program assumes that no ion pairs are formed and that histidines are not involved in a hydrogen bond. These data were used to construct the graph in Figure 10.

Protein Concentration Determination

The cell loading concentration for each sample in the solubility and mutant experiments was estimated using NONLIN (60). Residual graphs for data sets of the same concentration at different speeds were plotted. The concentration in fringes at the point where the data sets meet is called the "hinge point". This point is near the center of the cell and corresponds to the approximate cell loading concentration (77).

Data Analysis

Data from the ultracentrifuge experiments were edited using REEDIT (kindly provided by David Yphantis). Data sets

were truncated to avoid Weiner skewing at high fringe gradient (68) or noisy data at low fringe gradients. Each data set used for analysis included at least 50 data points, and up to 15 data sets were included in each experimental data analysis. The data were analyzed using nonlinear least squares technique (NONLIN) (60). This program provided fitting parameters and the limits for 65% confidence intervals. A model was considered to be correct if it provided an acceptable variance of fit and acceptable random systematic distribution of residuals. If more than one model was judged to fit these criteria, the model chosen was the simplest one (60).

The model which correctly fit to all self-associating experiments of IN in glycerol-free solutions was an ideal monomer-dimer-tetramer association. The model which best fit glycerol solutions of 8% (vol/vol) and higher was an ideal dimer-tetramer. The association models determined in these experiments were reversible according to previously established methods (60,69).

The model which was used for an ideal self-association is described by the equation (68):

Equation 1: $$Y(r) = \delta + e^{\ln A + \sigma\xi} + e^{N(\ln A + \sigma\xi) + \ln K}$$

where Y(r) is the fringe displacement (in fringes) at a given radius, δ is the baseline offset, A is the monomer activity at the arbitrary reference radius r_o, σ is the reduced molecular

weight, $\xi = (r^2-r_o^2)/2$, N is the stoichiometry of the self-association and K is the association constant.

The model returns 65% confidence intervals. These limits describe the precision of the fit for the model chosen. The confidence intervals which are given in the Results section are this measure of precision and are often not symmetric due to the fact that they are calculated on nonlinear data (60).

The NONLIN program returns a reduced apparent molecular weight. The reduced apparent molecular weight was converted into the apparent molecular weight using the following equation:

Equation 2:

$$M_z = \sigma_z \frac{RT}{(1-\bar{v}\rho)\omega^2}$$

where M_z is the z-average molecular weight, σ_z is the reduced apparent molecular weight, ρ is the density of the solution, R is the gas constant, T is the temperature of the experiment and ω is the angular velocity.

The association constants returned by NONLIN were in units of $\ln K_{fringes}$. These constants for best fit monomer-dimer and monomer-tetramer models were converted from reciprocal fringe displacement units to reciprocal molarity units by using the monomer molecular weight of 31,750 for IN in the following equation:

Equation 3:

$$K_a = K_n \times \left(\frac{M_m^{[n-1]}}{n}\right) \times \left(\frac{Y_t}{C_t}\right)^{n-1}$$

where K_a is the molar association constant, K_n is the fringe association constant, M_m is the monomer molecular weight, n is the stoichiometry of self-association and Y_t/C_t is the specific fringe displacement (3.333 fringes/mg/ml at 670 nm).

The reciprocal of the association constant is K_d, which is the dissociation constant. It is these values that are included in the tables of the Results section.

The association constants for the monomer-dimer equilibrium and monomer-tetramer equilibrium were returned by NONLIN with the 65% confidence interval. The dimer-tetramer association was calculated by the following equation:

Equation 4: $K_{2\rightarrow4} = K_4/(K_2)^2$

where K_4 is the association or dissociation constant for the monomer-tetramer species and K_2 is the corresponding constant for the monomer-dimer species. The confidence limits that were calculated with this value include the most extreme limits, therefore the lowest confidence limit of one species was used in the above equation with the highest confidence limit of the other species to determine the widest range of error possible. These values are included in the tables of the Results section.

The association or dissociation free energies of the oligomers were calculated using:

Equation 5: $\Delta G = -RT\ln(K_{eq})$

where ΔG is the free energy of association (or dissociation), R is the gas constant in joules/mol/degree K, T is the temperature in Kelvin and $\ln K_{eq}$ is the ln of the association (or dissociation) constant calculated in equation 3. It is the ΔG values of dissociation that appear in the tables and graphs of the Results section.

RESULTS

It has been shown that IN requires oligomerization for proper function (2,11,12,27,51,57,59). The stoichiometry of this association was not known, nor was it known if the association is reversible. The purpose of the first part of this study was to establish if RSV IN undergoes a reversible self-association and, if it does, to determine the best fit model for this association. Once this model of association was established, RSV IN was examined in a number of different solvent conditions to determine what effect, if any, a change in the physical environment had on the association properties of the protein.

Model of Self-Association

IN at a concentration of 1.0 mg/ml in a buffer containing 500 mM NaCl was examined in the ultracentrifuge as previously described. NONLIN analysis showed the z-average molecular weight to increase with cell loading concentrations to values in excess of 85,000. In comparison with the monomer molecular weight of 32,000, the data indicate that IN undergoes a mass action self association to oligomers larger than dimers. This association is reversible as judged by the criteria previously established (60,69). The data show that when the monomer molecular weight was held fixed during analysis, the best fit model was a reversible monomer-dimer-tetramer association.

The other types of models tested as well as the errors in these models are shown in Table 1.

Studies performed at the concentrations given in the Methods section confirmed the validity of the monomer-dimer-tetramer model in two ways: 1. When the monomer molecular weight of IN was considered a variable parameter, the value returned (31,700) was in close agreement with the value calculated from the cDNA sequence (31,864). 2. The monomer molecular weight and the stoichiometry of the dimer were held fixed at the values listed for 500 mM NaCl in Table 2, and all other parameters (including the dimer constant as well as the stoichiometry and association constant for larger species) were recalculated. The stoichiometry of the largest oligomer returned was 4.01 ± 0.18, and the dissociation constants were nearly identical to those listed in Table 2 at 500 mM (70).

Figures 5 and 6 show the concentrations and mole fractions calculated for each form of IN as a function of total protein concentration. At the lowest concentration examined (5 μM subunits), 23% of the total protein is in the form of dimer and less than 0.1% is in the form of a tetramer, whereas at the highest concentration (137 μM), 57% is a dimer and 18% is a tetramer (70). The best fit monomer-dimer-tetramer model established in this experiment was tested under different buffering conditions as described below. This model was the best-fit model in all glycerol-free experimental data analyzed.

TABLE 1

ATTEMPTED MODELS OF SELF-ASSOCIATION

Model	RMS fringes[a]	Comments
Monomer-Trimer	.020	Residual graphs showed downward hook indicating that a larger species is present. Returned sigma indicated a molecular weight less than monomer.
Monomer-Tetramer	.034	Residuals acceptable but returned sigma indicated a molecular weight less than monomer. RMS and residuals not as good as monomer-dimer-tetramer.
Dimer-Tetramer	N/A	NONLIN would not accept this model.
Monomer-Dimer-Trimer	.084	NONLIN would not return full statistics. Unacceptable model.
Monomer-Trimer-Hexamer	.045	NONLIN would not return full statistics. Unacceptable model.
Monomer-Dimer-Tetramer	.018	Good residuals. Returned sigma indicates a molecular weight in close agreement with cDNA calculated molecular weight (32,000).
Monomer-Dimer-Hexamer	.018	Returned sigma indicates a molecular weight in close agreement with cDNA calculated molecular weight (32,000). Residuals are pretty good but monomer-dimer-tetramer model looks a bit better and it is a simpler model.

[a]Root mean square of the variance of the fit.

Figure 5. Concentration of Multimeric Forms of RSV IN as a Function of Total Concentration. This graph was constructed from data listed in Table 2 for 500 mM NaCl. Dilutions were made from a sample of 1.0 mg/ml of IN equilibrated with 500 mM NaCl, 20 mM Tris (pH 8.4), and 2 mM BME by centrifugal gel chromatography. Loading concentrations were 1.0. 0.67, 0.44, 0.29 and 0.15 mg/ml. Rotor speeds were 20,000, 24,000, 28,000 and 36,000.

Concentration of RSV IN Multimers

Figure 5

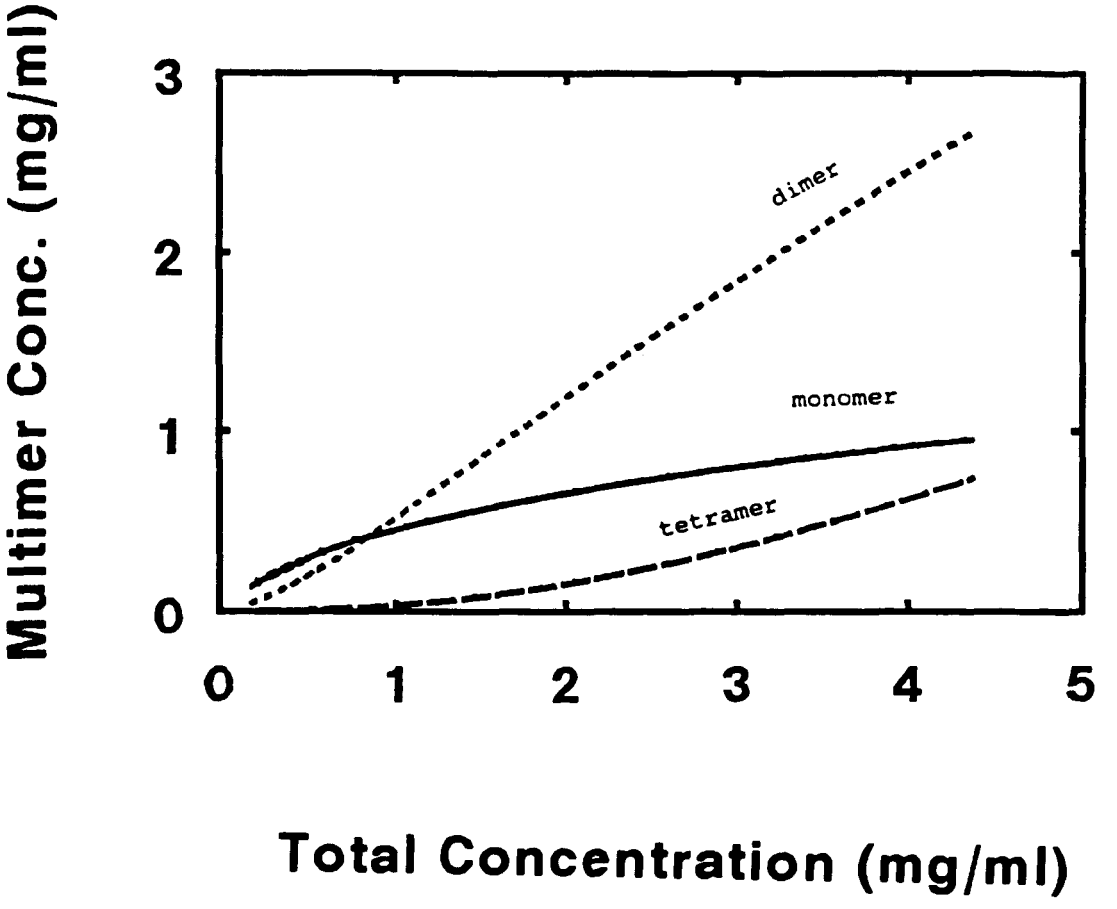
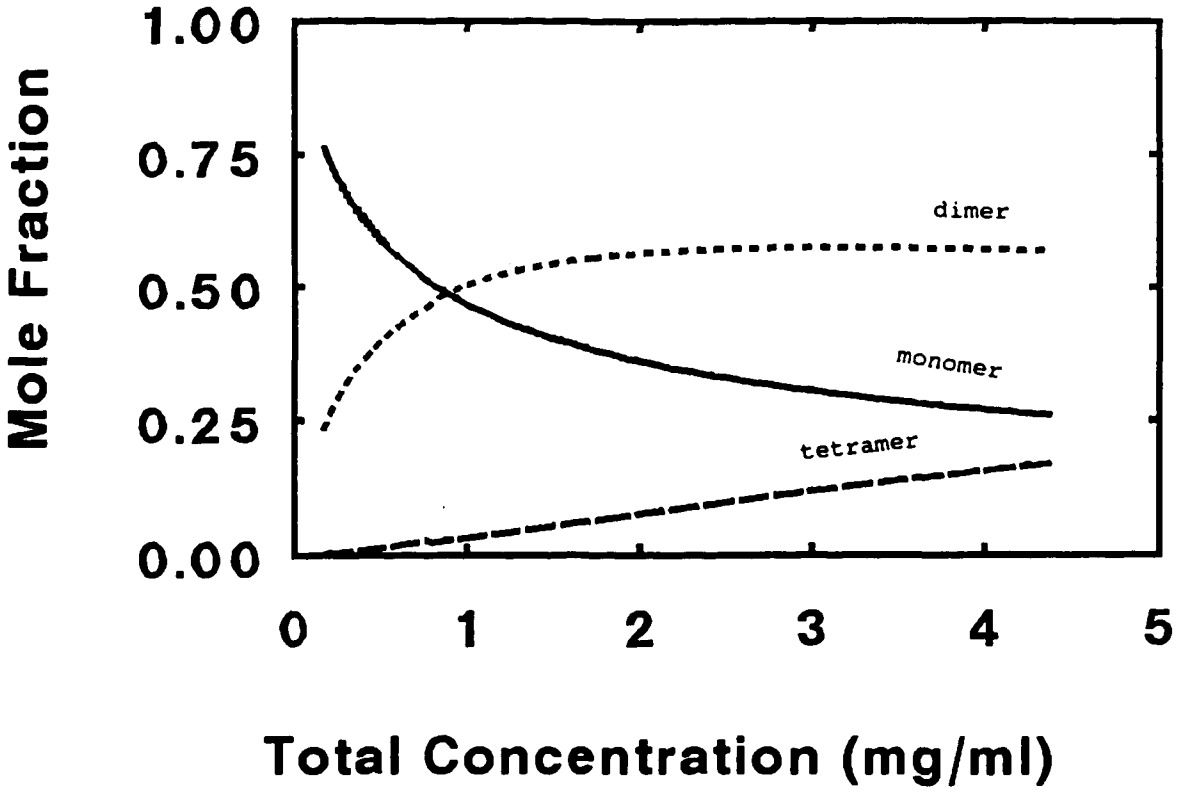


Figure 6. Mole Fraction of Multimeric Forms of RSV IN as a Function of Total Concentration. This graph was constructed from data listed in Table 2 for 500 mM NaCl. Dilutions were made from a sample of 1.0 mg/ml of IN equilibrated with 500 mM NaCl, 20 mM Tris (pH 8.4), and 2 mM BME by centrifugal gel chromatography. Loading concentrations were 1.0. 0.67, 0.44, 0.29 and 0.15 mg/ml. Rotor speeds were 20,000, 24,000, 28,000 and 36,000.

Mole Fraction of RSV IN Multimers

Figure 6



Effect of Ionic Strength on Solubility of IN

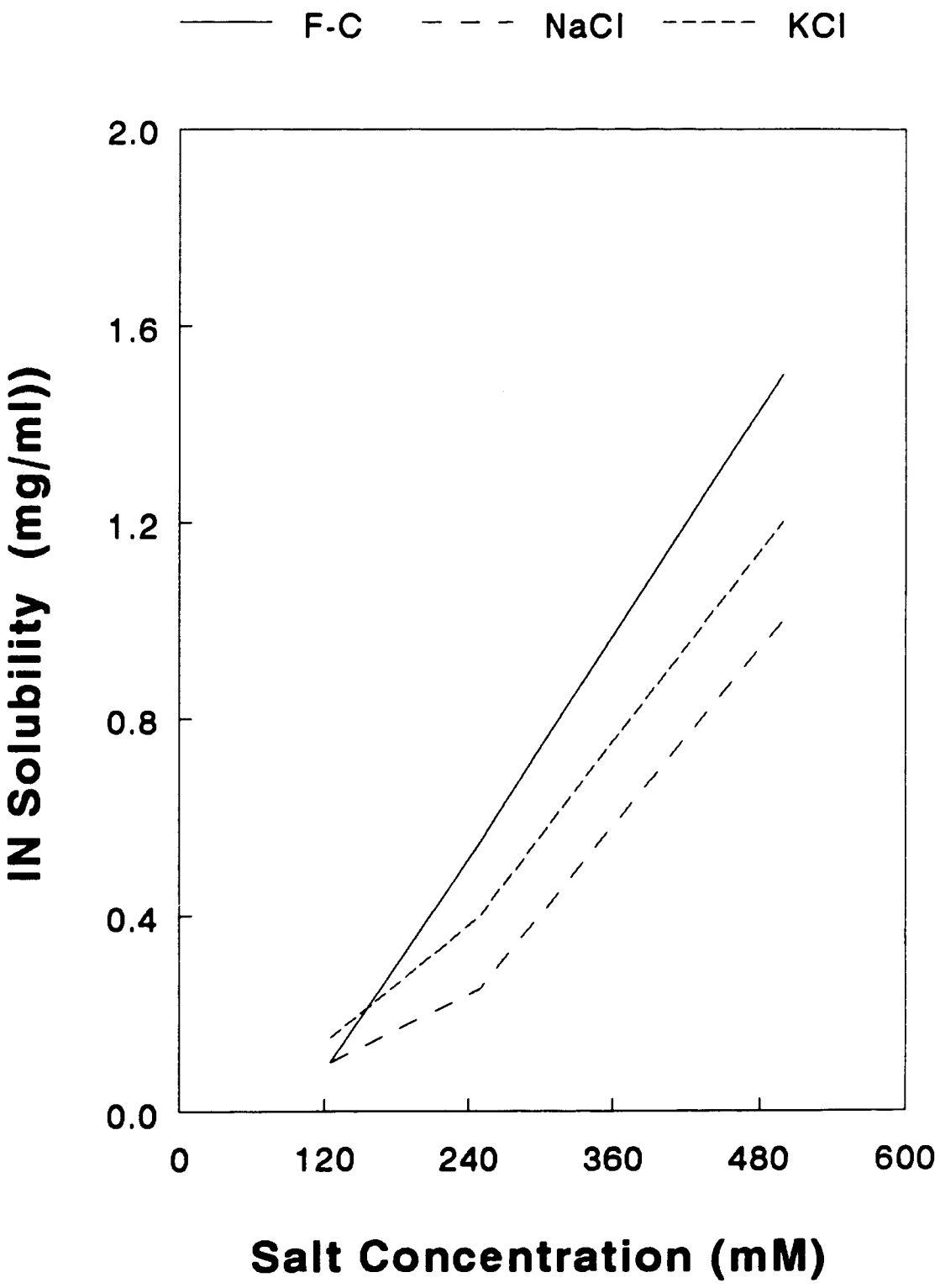
Experiments performed as part of the study described above showed that IN requires high NaCl concentration for solubility. It had been observed that at the protein concentration needed for experiments in the ultracentrifuge, IN required NaCl concentrations greater than physiological (145 mM). However, the exact solubility characteristics were unknown. This problem was discussed with the laboratory staff of Dr. Skalka who supplied the protein used in these experiments. Figure 7 contains information given by Skalka's lab based on their observations of the solubility of IN (data labelled F-C). Because part of the purpose of these experiments is to test the association properties of IN in different solvent conditions, it is important to study the solubility requirements of this protein.

The first set of experiments was designed to test the solubility of IN in NaCl. In one experiment, IN at a concentration of 1.0 mg/ml was equilibrated as described in buffers containing NaCl concentrations of 62.5, 125, 250 and 500 mM, while in another experiment 1.5 mg/ml of IN was equilibrated in buffers of 150, 250, 550, 750 and 1000 mM NaCl. Equilibrium sedimentation studies showed no fringe displacement in the 62.5 mM NaCl cell, indicating that the IN concentration was less than 0.05 mg/ml. The cell containing 125 mM NaCl showed a concentration of 0.1 mg/ml. While this is not enough protein in the solution to provide sufficient

Figure 7. Solubility of RSV IN. This graph is the solubility of IN in mg/ml as a function of salt concentration. The solid line labelled F-C in the legend is data which was obtained from Dr. Skalka's laboratory at Fox Chase and is based on their observations when preparing purified IN. The smaller broken line is the solubility of IN in 20 mM Tris (pH 8.4), 2 mM BME and KCl concentrations of 125, 250 and 500 mM. The larger broken line is the same concentration of buffer with NaCl used as the salt in the place of KCl. The protein concentration in each buffer was determined by using the "hinge point" method described in the Materials and Methods section. Buffer exchange was performed by centrifugal gel chromatography.

RSV IN Solubility

Figure 7



data sets for NONLIN analysis, it does show that the IN is becoming soluble at this concentration. It was in the cell beginning with 250 mM of NaCl that sufficient data sets for NONLIN analysis could be collected. The protein concentration in this cell was 0.3 mg/ml. The solubility characteristics of IN in NaCl as determined in this experiment are shown in Figure 7.

The experiments described above were performed in NaCl because many experiments have been reported in NaCl-containing buffers (10,13,35,39,50,52,54,55,57,58). Other published work includes KCl as the salt component of the buffers used to measure IN reactions (27,28,34,47,48). RSV IN is an intracellular protein and the intracellular concentration of Na^+ is 5 - 15 mM while that of K^+ is 140 mM (71). It was important to know if there is a difference in the behavior of IN in buffers containing NaCl from those buffers which contain KCl. A series of experiments was performed to determine the effect of KCl ionic strength on RSV IN solubility.

IN at a concentration of 1.2 mg/ml was equilibrated as described in buffers containing KCl concentrations of 75, 125, 250, 500, 750 and 1000 mM KCl. Equilibrium ultracentrifugation results showed that the protein concentration was less than 0.05 mg/ml, but was present at a concentration of 0.15 mg/ml in the 125 mM KCl solution. The results of the solubility of IN in the KCl solutions is given in Figure 7. A review of this graph shows that IN is somewhat

more soluble in KCl than in NaCl.

Effect of Ionic Strength on Self-Association

The experiments which provided sufficient data sets for NONLIN from the solubility experiments described above were analyzed to determine the effect of ionic strength on the self-association of IN. The differences in the two cations, Na^+ and K^+ were also studied as a part of this experiment. Sufficient data sets for NONLIN analysis were therefore obtained in the experiments using 1.0 mg/ml of protein in 250 and 500 mM NaCl, and 1.5 mg/ml of protein in 550, 750 and 1000 mM NaCl. The data from these experiments were analyzed as described to determine the stoichiometries and K_d values of IN under conditions of increasing NaCl ionic strength. The results of these experiments are given in Table 2. Figure 8 is a graph of the dependence of the ΔG values on the NaCl concentration. The monomer-dimer association increases with ionic strength to 550 mM NaCl, then appears to remain fairly constant to 1000 mM NaCl. The monomer-tetramer association increases with increasing NaCl concentration to 750 mM. There is no significant difference in the monomer-tetramer association of the 750 and 1000 mM sample (Table 2). The dimer-tetramer association decreases from 250 to 500 mM NaCl. There is no statistical difference in the values of the dimer-tetramer association at NaCl concentrations equal to or greater than 500 mM. These data indicate that solubility and

TABLE 2

EFFECT OF NaCl IONIC STRENGTH ON RSV INTEGRASE AT pH 8.4

NaCl mM[a]	$K_{d\ 2\text{-}1}$ μM[b]	$\Delta G_{\ 2\text{-}1}$ kJ/mol[c]	$K_{d\ 4\text{-}1}$ M^3x10^{-144}	$\Delta G_{\ 4\text{-}1}$ kJ/mol[d]	$\Delta G_{\ 4\text{-}2}$ kJ/mol[e]	RMS fringes[g]
250	157.8 (133.1; 188.9)	21.5 (21.1; 22.0)	54.4 (40.7; 73.5)	69.5 (68.7; 70.2)	26.5 (24.7; 28.0)	.009
500	5.7 (4.7; 6.8)	29.7 (29.3; 30.2)	0.4 (0.3; 0.5)	81.6 (80.9; 81.9)	22.2 (20.5; 23.3)	.018
550	1.0 (0.8; 1.3)	34.0 (33.3: 34.5)	0.03 (0.02; 0.04)	87.9 (87.2; 88.9)	19.9 (18.2; 22.3)	.014
750	0.5 (0.4; 0.8)	35.5 (34.5; 36.2)	.006 (.004; .012)	91.9 (90.2; 92.9)	20.9 (17.8; 23.9)	.015
1000	1.0 (0.7; 1.2)	34.1 (33.5; 34.9)	.012 (.006; .025)	90.2 (88.4; 91.9)	22.0 (18.6; 24.9)	.014

[a]NaCl concentration in 20 mM Tris(pH 8.4) and 2 mM BME.

[b]Monomer formation constant from dimer calculated from ln K returned by NONLIN holding **θ** fixed at the monomer molecular weight. All data fit as an ideal monomer-dimer-tetramer. Values in parentheses are the 65% confidence interval.

[c]Free energy of dissociation of dimer to monomer calculated from ln K returned by NONLIN holding **θ** fixed at the monomer molecular weight. Values in parentheses are the 65% confidence interval.

[d]Monomer formation constant from tetramer calculated as above (note b).

[e]Free energy of dissociation of tetramer to monomer calculated as above (note c).

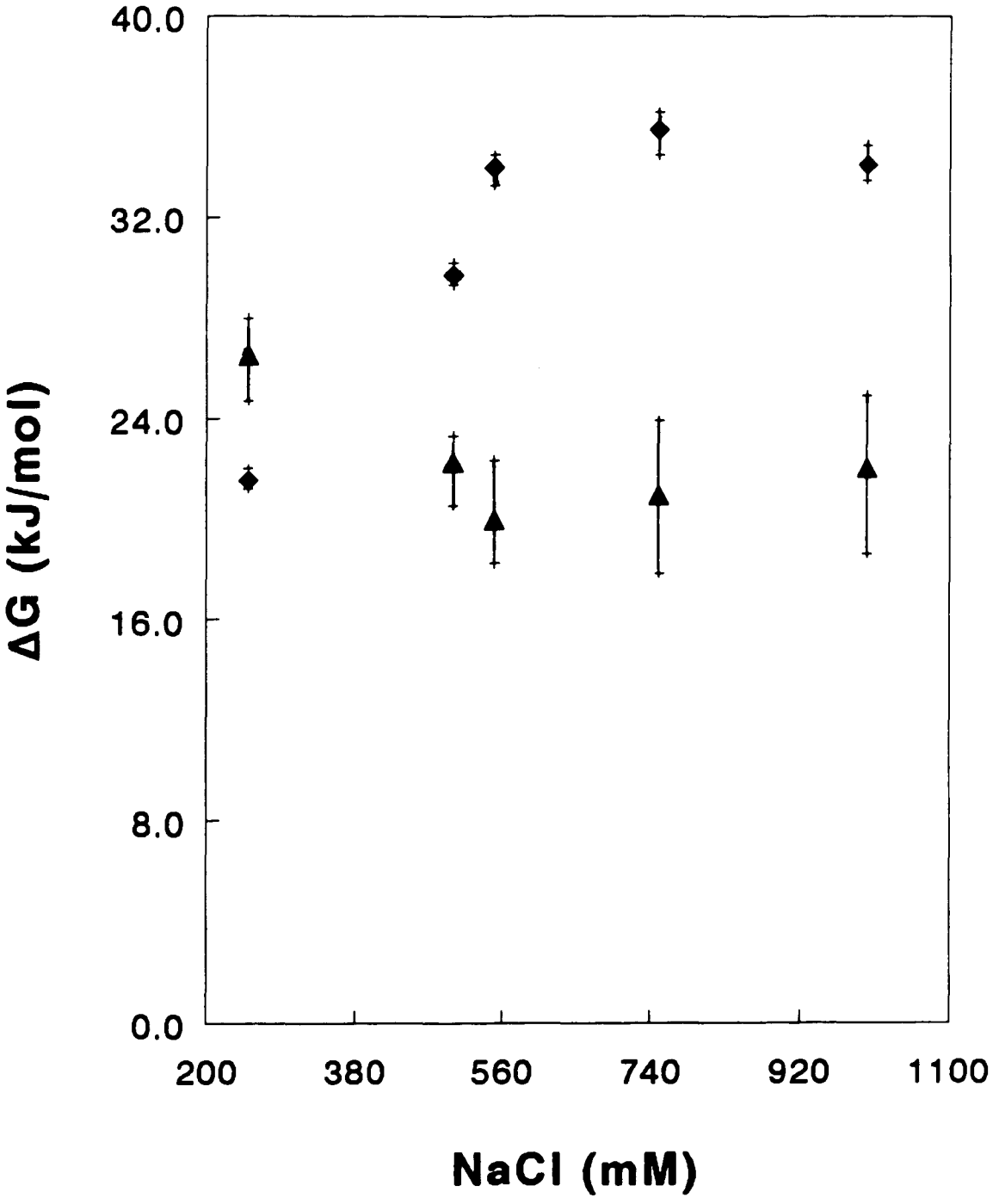
[f]Free energy of dissociation of tetramer to dimer calculated as described in "Materials and Methods".

[g]Root mean square of the variance of the fit.

Figure 8. Effect of NaCl Ionic Strength at pH 8.4 on the Extent of Association of RSV IN. Free energy of dissociation of IN as a function of NaCl ionic strength. The solid diamonds represent the free energy of dissociation of dimer to monomer. The solid triangles represent the free energy of dissociation of tetramer to dimer. IN was equilibrated by centrifugal gel chromatography into buffers of 20 mM Tris (8.4), 2 mM BME the NaCl concentration indicated. Protein loading concentrations were 0.3 - 1.5 mg/ml. K_d values were calculated from LnK returned by NONLIN from simultaneous fit to data acquired at rotor speeds of 20,000, 24,000, 28,000 and 36,000. Free energy values were calculated as described in the Materials and Methods section. Error bars represent the 65% confidence level.

NaCl Ionic Strength pH 8.4

Figure 8



self-association are not positively correlated and that the monomer-dimer and dimer-tetramer equilibrium have different ionic dependencies.

NONLIN analysis of data from the KCl buffers indicated the same model of association as the NaCl buffers. The results of these experiments are given in Table 3. Figure 9 is a graph of the dependence of the ΔG values on KCl concentration. A comparison of the values of Tables 2 and 3 demonstrates that both the monomer-dimer and the monomer-tetramer associations are slightly increased in NaCl solutions compared to KCl solutions of equal ionic strength. There is no statistical difference in the dimer-tetramer association in any of the data sets when comparing a buffer containing NaCl with a buffer containing an equal concentration of KCl. It can be concluded that, under the conditions studied, IN monomer-dimer and monomer-tetramer associations are slightly increased in NaCl solutions relative to KCl solutions, but that the dimer-tetramer association is the same in both types of salt solutions.

Effect of Lowering the pH on the Ionic Strength Dependency of IN Solubility

A graph of the predicted charge of IN as a function of pH was constructed using VBAR (Figure 10). It is calculated that at a pH of 8.4 the protein has a strong positive charge which increases with decreasing pH. An experiment then was performed to determine the effect of a decreased pH on the solubility

TABLE 3

EFFECT OF KCL IONIC STRENGTH ON RSV INTEGRASE AT pH 8.4

KCl mM*	$K_{d\ 2\text{-}1}$ μM	$\Delta G_{\ 2\text{-}1}$ kJ/mol	$K_{d\ 4\text{-}1}$ M^3x10^{-14}	$\Delta G_{\ 4\text{-}1}$ kJ/mol	$\Delta G_{\ 4\text{-}2}$ kJ/mol	RMS fringes
250	270.8 (215.1; 344.2)	20.2 (19.6; 20.8)	76.5 (53.9; 110.7)	68.6 (67.7; 69.4)	28.2 (26.1; 30.1)	.010
500	8.9 (7.4; 10.5)	28.6 (28.2; 29.0)	10.6 (4.4; 123.8)	73.5 (67.4; 75.6)	16.3 (9.4; 19.2)	.014
750	5.3 (4.1; 6.5)	29.9 (29.3; 30.5)	1.5 (1.1; 2.1)	78.3 (77.5; 79.1)	18.5 (16.5; 20.5)	.014
1000	3.6 (2.6; 4.9)	30.8 (29.9; 31.6)	0.6 (0.4; 0.8)	80.6 (79.8; 81.6)	19.0 (16.6; 21.8)	.017

*In 20 mM Tris (pH 8.4), 2 mM BME and the indicated concentration of KCl
All other constants as defined in Table 2.

Figure 9. Effect of KCl Ionic Strength at pH 8.4 on the Extent of Association of RSV IN. Free energy of dissociation of IN as a function of KCl ionic strength. The solid diamonds represent the free energy of dissociation of dimer to monomer. The solid triangles represent the free energy of dissociation of tetramer to dimer. IN at 1.2 mg/ml was equilibrated by centrifugal gel chromatography into buffers of 20 mM Tris (8.4), 2 mM BME and the KCl concentration indicated. Protein loading concentrations were 0.3 - 1.2 mg/ml. K_d values were calculated from LnK returned by NONLIN from simultaneous fit to data acquired at rotor speeds of 20,000, 24,000, 28,000 and 36,000. Free energy values were calculated as described in the Materials and Methods section. Error bars represent the 65% confidence level.

KCl Ionic Strength pH 8.4

Figure 9

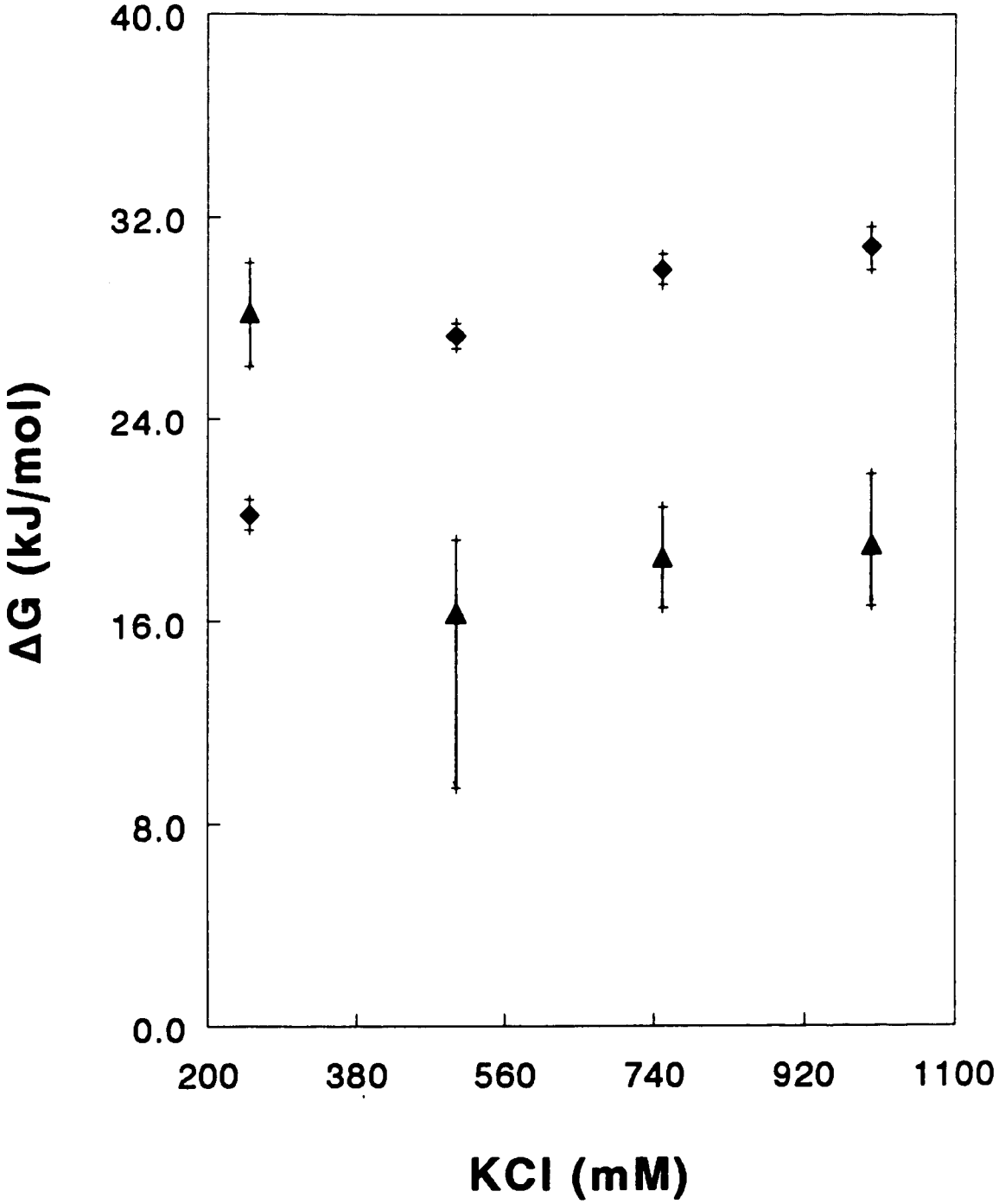
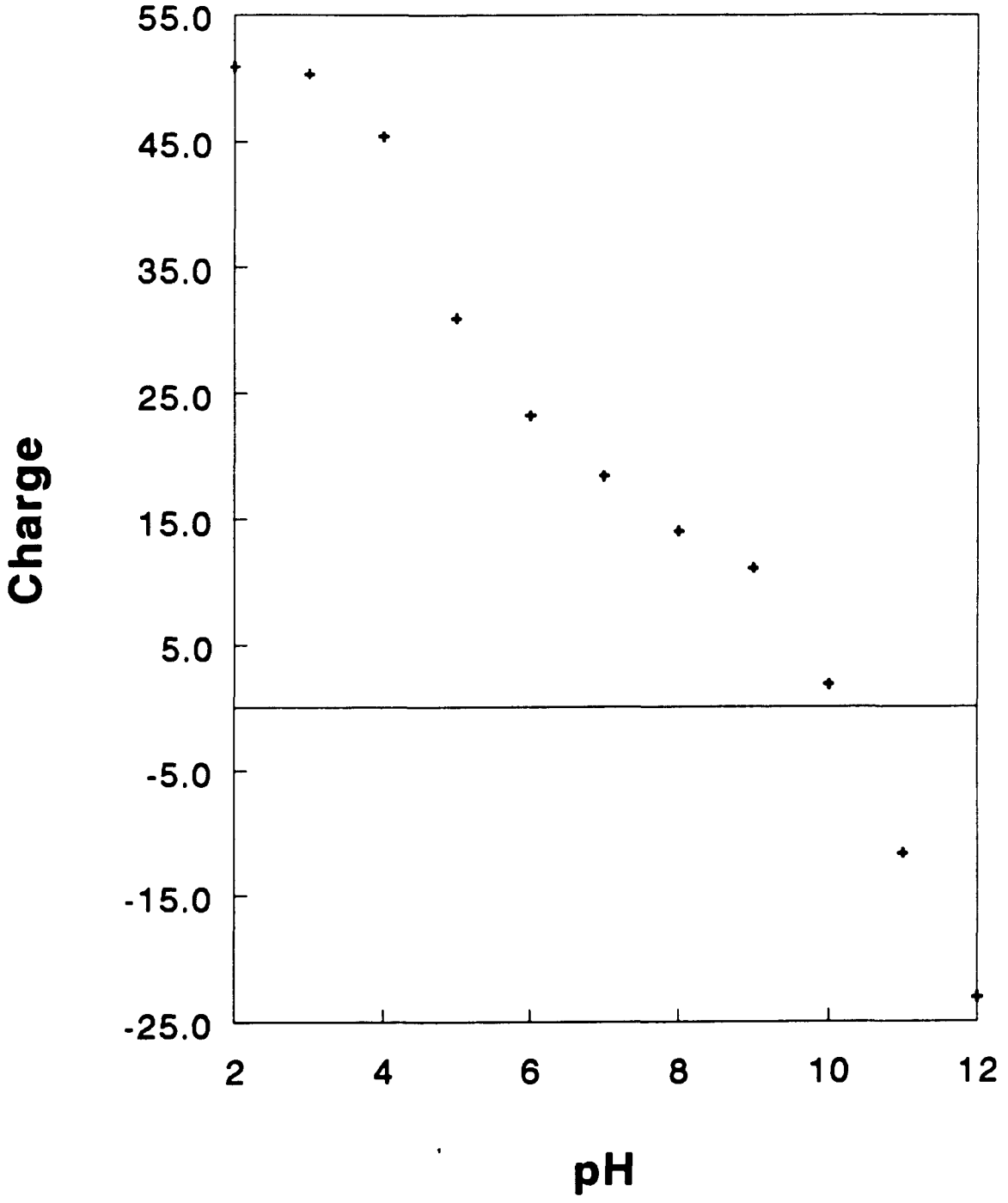


Figure 10. Charge on RSV IN. The charge on the IN molecule was calculated using the VBAR program. This calculated charge is based on the pKa values of the individual amino acids as described in the Materials and Methods section.

Charge on RSV Integrase

Figure 10



and association properties of IN in increasing ionic strength solutions of NaCl and KCl. The purpose of this experiment was to determine the effect of a pH change on the stoichiometry and association constants of IN. The results of these experiments will also be used by a group that is using NMR techniques to study IN. RSV IN at a concentration of 1.2 mg/ml was equilibrated as described in buffers containing 125, 250 and 750 mM NaCl at a pH of 6.1. The results of the NONLIN analysis of this experiment are given in Table 4. The values of the ΔG dependence on NaCl concentration at pH 6.1 are given in Figure 11.

The solubility of IN increases in NaCl when the pH is lowered from 8.4 to 6.1 as demonstrated by the fact that the protein remains soluble in the 125 mM NaCl sample used in this experiment. The data show that all NaCl concentrations studied have a monomer-dimer association which strengthens with increasing salt concentrations to 500 mM. The corresponding dimer-tetramer formation becomes weaker. The reason for the increase in the ΔG of tetramer to dimer of the 750 mM NaCl sample is unknown. The ΔG values of the monomer-dimer association in Table 4 in comparison with Table 2 are significantly higher in the 250 mM specimen at pH 6.1, slightly higher in the 500 mM sample at pH 6.1 and significantly lower in the 750 mM sample at pH 6.1. It is therefore not possible to generalize what the overall effect of decreasing the pH from 8.4 to 6.1 is on the association

TABLE 4

EFFECT OF NaCl IONIC STRENGTH ON RSV INTEGRASE AT pH 6.1

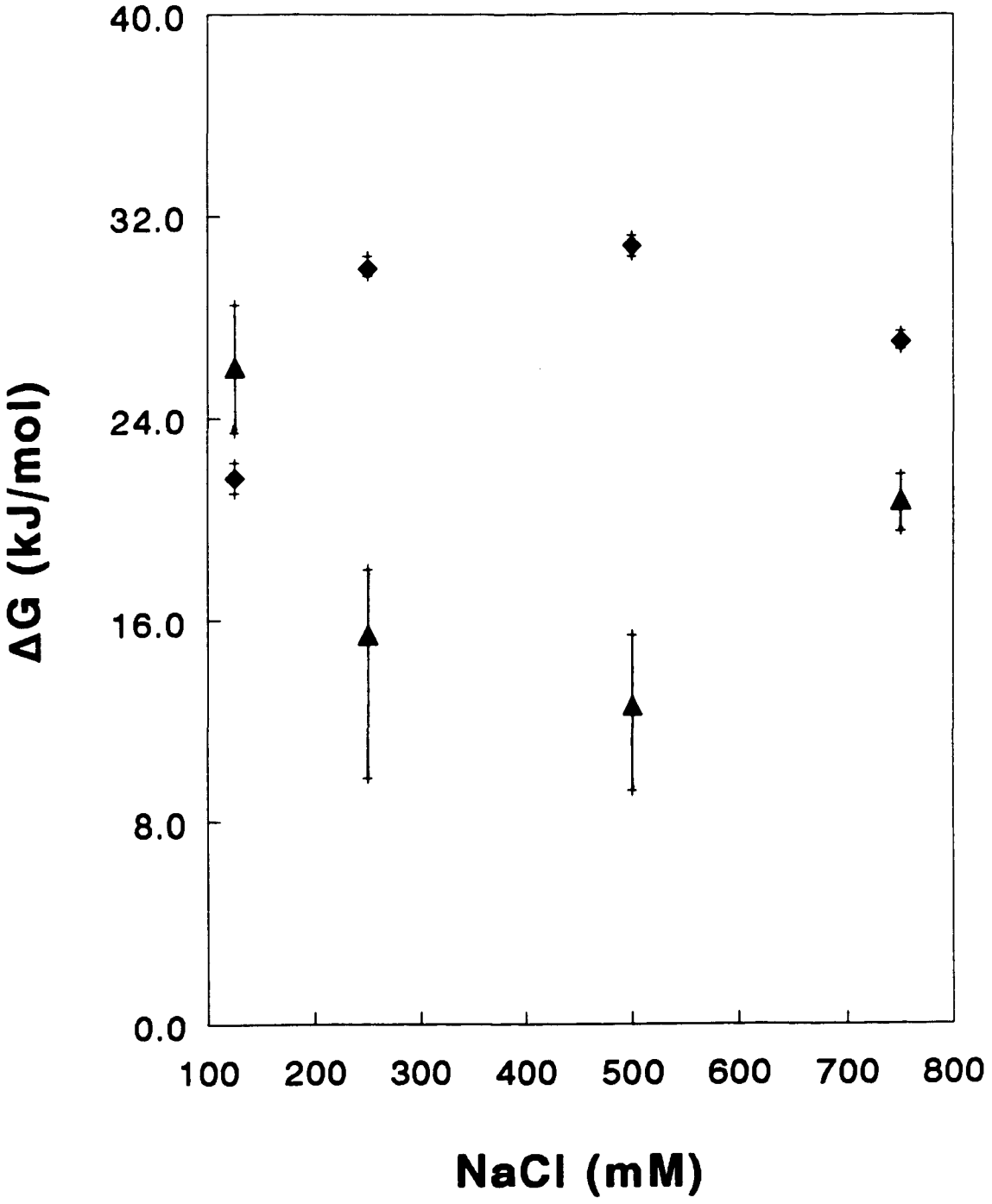
NaCl mM[a]	$K_{d\ 2-1}$ μM	ΔG_{2-1} kJ/mol	$K_{d\ 4-1}$ M^3x10^{-14}	ΔG_{4-1} kJ/mol	ΔG_{4-2} kJ/mol	RMS fringes
125	154.7 (121.7; 196.6)	21.6 (21.0; 22.2)	59.0 (35.8; 104.3)	69.2 (67.8; 70.5)	26.0 (23.4; 28.5)	.011
250	5.2 (4.3; 6.0)	29.9 (29.6; 30.4)	5.4 (2.2; 18.7)	75.2 (70.5; 77.4)	15.4 (9.7; 18.0)	.010
500	3.7 (3.1; 4.3)	30.8 (30.4; 31.2)	7.9 (3.5; 23.3)	74.2 (71.6; 76.2)	12.6 (9.2; 15.4)	.014
750	16.9 (14.5; 19.6)	27.0 (26.7; 27.4)	6.5 (5.6; 7.5)	74.7 (74.3; 75.1)	20.7 (19.5; 21.7)	.010

[a]Buffer is 50 mM 2[N-Morpholino]ethanesulfonic acid (MES)(pH 6.1), 2 mM BME and the indicated concentration of NaCl. All other constants defined in Table 2.

Figure 11. Effect of NaCl Ionic Strength at pH 6.1 on the Extent of Association of RSV IN. Free energy of dissociation of IN as a function of NaCl ionic strength. The solid diamonds represent the free energy of dissociation of dimer to monomer. The solid triangles represent the free energy of dissociation of tetramer to dimer. IN at 1.2 mg/ml was equilibrated by centrifugal gel chromatography into buffers of 20 mM Tris (6.1), 2 mM BME and the NaCl concentration indicated. Protein loading concentrations were 0.3 - 1.2 mg/ml. K_d values were calculated from LnK returned by NONLIN from simultaneous fit to data acquired at rotor speeds of 20,000, 24,000, 28,000 and 36,000. Free energy values were calculated as described in the Materials and Methods section. Error bars represent the 65% confidence level.

NaCl Ionic Strength pH 6.1

Figure 11



strength under the conditions studied. It is possible to conclude, however, that the model of association in all specimens studied is a monomer-dimer-tetramer.

IN at a concentration of 1.2 mg/ml was equilibrated as described in buffers containing 75, 125, 250, 500 and 750 mM KCl. Equilibrium ultracentrifugation results showed that the protein had precipitated in the 75 mM solution but had remained soluble in the remaining concentrations of KCl.

The data from these experiments were analyzed by NONLIN and used to determine the stoichiometries and K_d values of IN under conditions of increasing KCl ionic strength at a pH of 6.1. The results of these experiments are given in Table 5. The values of the ΔG dependence on KCl concentration at pH 6.1 are given in Figure 12. The data show that the monomer-dimer association becomes stronger with increasing ionic strength. The monomer-dimer ΔG values for the KCl experiments in Table 5 in comparison with Table 3 are significantly increased in the 250 mM solution at pH 6.1, slightly decreased in the 500 mM sample at pH 6.1 and significantly decreased in the 750 mM sample at pH 6.1. This parallels the comparison made at the same molar concentrations in NaCl. As with the NaCl data, it is not possible to conclude what the overall effect of decreasing the pH from 8.4 to 6.1 is on IN association energies in KCl under the conditions studied. There is no significant difference between the ΔG values of both the monomer-dimer and dimer-tetramer association in 750 mM NaCl

TABLE 5

EFFECT OF KCl IONIC STRENGTH ON RSV INTEGRASE AT pH 6.1

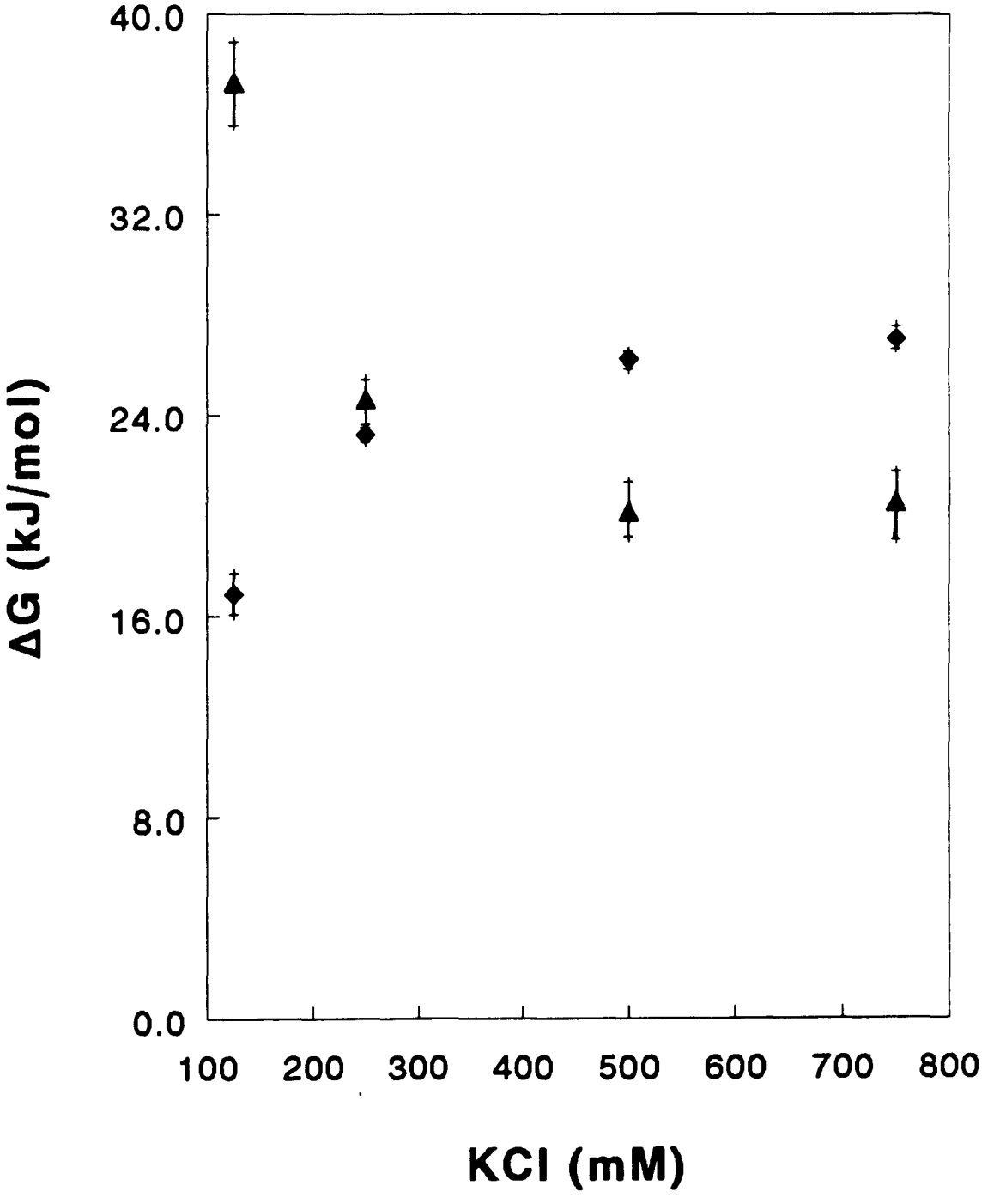
KCl mM*	$K_{d\ 2-1}$ μM	$\Delta G_{\ 2-1}$ kJ/mol	$K_{d\ 4-1}$ M^3x10^{-14}	$\Delta G_{\ 4-1}$ kJ/mol	$\Delta G_{\ 4-2}$ kJ/mol	RMS fringes
125	1023.8 (758.4; 1424.0)	16.9 (16.1; 17.7)	28.4 (27.6; 29.6)	71.1 (71.0; 71.1)	37.3 (35.6; 38.9)	.006
250	78.4 (68.2; 89.2)	23.2 (22.9; 23.6)	29.6 (27.0; 32.7)	71.0 (70.7; 71.2)	24.6 (23.5; 25.4)	.007
500	23.8 (20.6; 27.5)	26.2 (25.8; 26.5)	15.9 (13.4; 18.9)	72.5 (72.1; 72.9)	20.1 (19.1; 21.3)	.010
750	17.0 (14.2; 20.2)	27.0 (26.6; 27.5)	7.1 (5.9; 8.6)	74.5 (74.0; 74.9)	20.5 (19.0; 21.7)	.014

*Buffer is 50 mM MES(pH 6.1), 2 mM 2-mercaptoethanol and the indicated concentration of KCl. All other constants as defined in Table 2.

Figure 12. Effect of KCl Ionic Strength at pH 6.1 on the Extent of Association of RSV IN. Free energy of dissociation of IN as a function of KCl ionic strength. The solid diamonds represent the free energy of dissociation of dimer to monomer. The solid triangles represent the free energy of dissociation of tetramer to dimer. IN at 1.2 mg/ml was equilibrated by centrifugal gel chromatography into buffers of 20 mM Tris (6.1), 2 mM BME and the KCl concentration indicated. Protein loading concentrations were 0.3 - 1.2 mg/ml. K_d values were calculated from LnK returned by NONLIN from simultaneous fit to data acquired at rotor speeds of 20,000, 24,000, 28,000 and 36,000. Free energy values were calculated as described in the Materials and Methods section. Error bars represent the 65% confidence level.

KCl Ionic Strength pH 6.1

Figure 12



and KCl at this pH.

Effect of pH

Because no conclusion could be reached in the previous experiments concerning the effect of lowering the pH on the free energy of association of IN, it was decided to do a series of experiments on the effect of pH on RSV IN. The previous studies showed that the experiments performed in 500 mM salt resulted in sufficient rumber of data sets for use in NONLIN. The 500 mM KCl concentration buffer was therefore chosen to use for this set of experiments.

Figure 10 shows the predicted total charge of IN as a function of pH. According to this prediction IN becomes a positively charged molecule at a pH close to 10. Decreasing the pH increases this positive charge, so that at a physiological pH of 7.5 the charge on the molecule is about +20. An experiment was conducted to determine how a pH range of 6.1 to 8.4 would affect the association properties of IN. This range is chosen because it will include the pH extremes of buffers used in all experiments listed in the references. The corresponding predicted total charge on the molecule over this range would be from +13 to +23.

The experiment was performed at a pH of 6.1, 7.5, 8.0 and 8.4 in buffers as previously described. The results of these experiments are given in Table 6. Figure 13 is a graph of the dependence of the free energies on the pH at the conditions studied.

TABLE 6

EFFECT OF pH ON RSV INTEGRASE

pH[a]	$K_{d\ 2-1}$ μM	ΔG_{2-1} kJ/mol	$K_{d\ 4-1}$ M^3x10^{-14}	ΔG_{4-1} kJ/mol	ΔG_{4-2} kJ/mol	RMS fringes
6.1	3.7 (3.1; 4.3)	30.8 (30.4; 31.2)	7.9 (3.5; 23.3)	74.2 (71.6; 76.2)	12.6 (9.2; 15.4)	.014
7.5	66.1 (56.9; 77.6)	23.7 (23.3; 24.0)	23.0 (20.6; 25.7)	71.6 (71.3; 71.8)	24.2 (23.3; 25.2)	.012
Repeat[b]	74.5 (62.2; 90.1)	23.4 (22.9; 23.8)	27.3 (23.0; 32.4)	71.2 (70.7; 71.6)	24.4 (23.1; 25.8)	.015
8.0	14.5 (11.5; 18.7)	27.4 (26.8; 28.0)	2.7 (2.1; 3.4)	76.8 (76.2; 77.5)	22.0 (20.2; 23.9)	.017
8.4	15.1 (12.4; 18.3)	27.3 (26.8; 27.8)	3.0 (2.4; 3.6)	76.6 (76.1; 77.1)	22.0 (20.5; 23.5)	.011

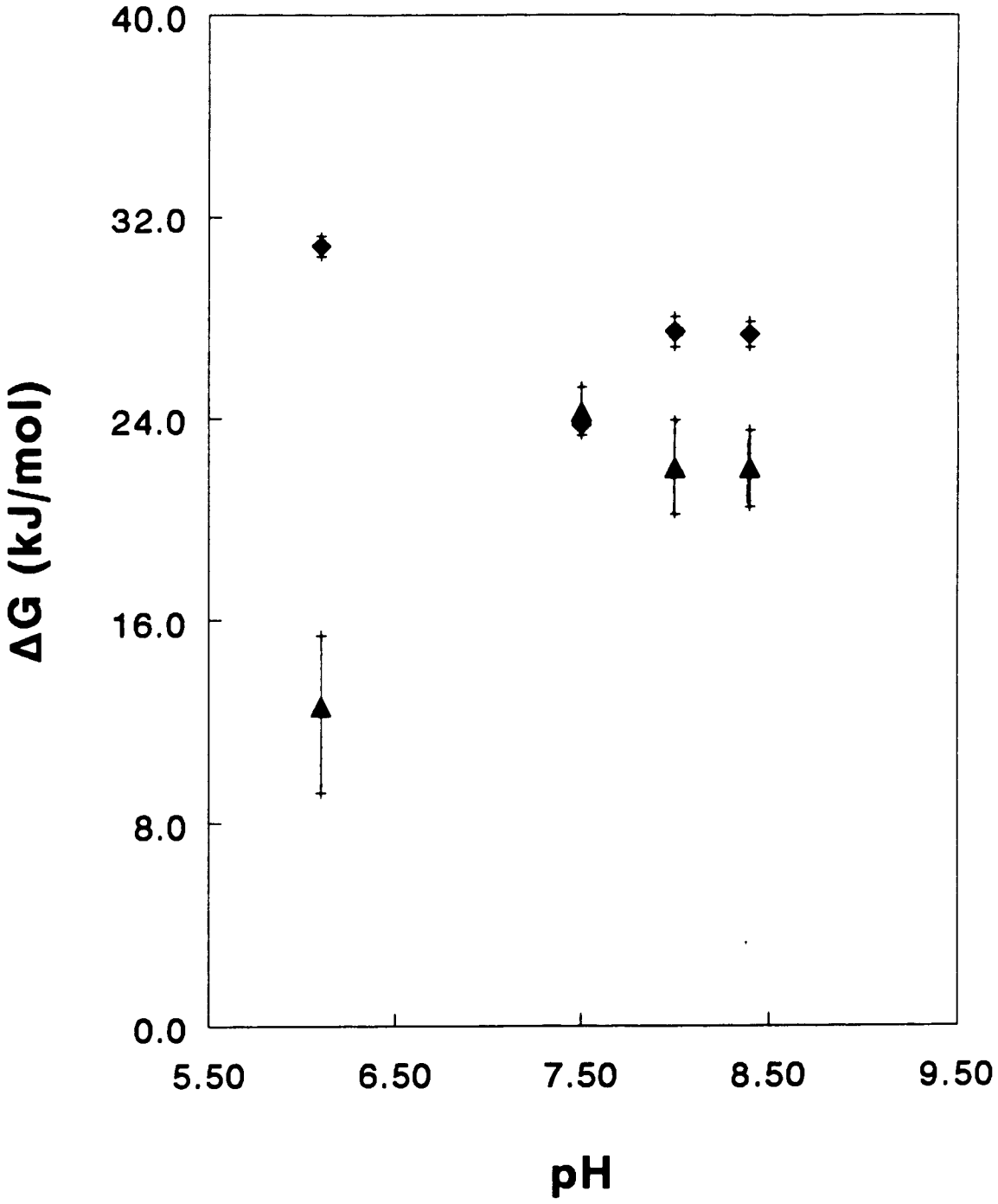
[a]In 500 mM KCl, 2 mM BME and the following buffer at the indicated pH: 20 mM Tris(8.4); 20 mM Tris(7.5); 50 mM HEPES(8.0); 50 mM MES(6.1). All other constants as defined in Table 2.

[b]Results of the repeat experiment are shown for comparison. The results of the first experiment are used in Figure 13 and in the discussion.

Figure 13. Effect of pH Change on the Extent of Association of RSV IN. The solid diamonds represent the free energy of dissociation of dimer to monomer. The solid triangles represent the free energy of dissociation of tetramer to dimer. IN at 1.2 mg/ml was equilibrated by centrifugal gel chromatography into 500 mM KCl, 2 mM BME and the following buffer at the indicated pH: 20 mM Tris (8.4), 20 mM Tris (7.5), 50 mM HEPES (8.0), 50 mM MES (6.1) Protein loading concentrations were 0.3 - 1.2 mg/ml. K_d values were calculated from LnK returned by NONLIN from simultaneous fit to data acquired at rotor speeds of 20,000, 24,000, 28,000 and 36,000. Free energy values were calculated as described in the Materials and Methods section. Error bars represent the 65% confidence level.

Effect of pH on IN

Figure 13



The data show that there is no significant difference in any of the dissociation constants of the specimens at a pH of 8.4 or 8.0. At a pH of 6.1 the dimer formation is strengthened and the tetramer formation is weakened.

Effect of Glycerol Concentration

As with many other proteins, glycerol is used to stabilize solutions of IN. All of the experiments discussed so far have been performed in the absence of glycerol. Published studies of IN include work that is performed in the presence of glycerol (10,13,34,35,39,41,50,52,55,57,58). It has been reported that the presence of glycerol creates additional products when studying the cleaving reaction of IN (50). It therefore becomes important to know the effect of glycerol on the association properties of IN. Glycerol studies were performed as described in the methods section on dilutions of IN at an initial concentration of 1.2 mg/ml equilibrated in buffers containing either 4, 8, 16, 20 or 40% glycerol. The results of these experiments are given in Table 7. The dependence of the free energy of dissociation on glycerol concentration is shown in Figure 14.

The data for 4% (vol/vol) glycerol fit the same model as the 500 mM NaCl experiment with 0% glycerol that is described in the section on ionic strength. This best fit association was monomer-dimer-tetramer, however only the dimer-tetramer values are included in Table 7 for comparison. Increasing the glycerol to 8, 16, 20 and 40% (vol/vol) produced a significant

TABLE 7

EFFECT OF GLYCEROL CONCENTRATION

Percent Glycerol (vol/vol)[a]	Best-fit Model[b]	K_{4-2} μM[c]	ΔG_{4-2} kJ/mol[d]	RMS[e] fringes
0	monomer-dimer-tetramer	125.0 (74.3; 237.7)	22.2 (20.5; 23.3)	.018
4	monomer-dimer-tetramer	299.4 (145.4; 509.1)	20.0 (18.6; 21.7)	.010
8	dimer-tetramer	413.4 (306.2; 575.0)	19.2 (18.4; 19.9)	.016
16	dimer-tetramer	100.9 (84.3; 122.0)	22.6 (22.2; 23.1)	.020
20	dimer-tetramer	111.5 (96.0; 129.6)	22.4 (22.0; 22.8)	.017
40	dimer-tetramer	28.3 (23.2; 34.2)	25.8 (25.3; 26.2)	.006

[a]In 500 mM NaCl, 20 mM Tris(8.4), and 2-mM BME.

[b]Monomer molecular weight held fixed for 0 and 4% data sets. Dimer molecular weight held fixed for all other data sets.

[c]Dimer formation constant from tetramer calculated from ln K returned by NONLIN holding θ fixed at the dimer molecular weight for 8,16,20 and 40% data sets. The 0 and 4% data sets were calculated from the ideal monomer-dimer-tetramer best fit as described in materials and methods. This calculation assumes the greatest margin of error possible. Values in parentheses are the 65% confidence interval.

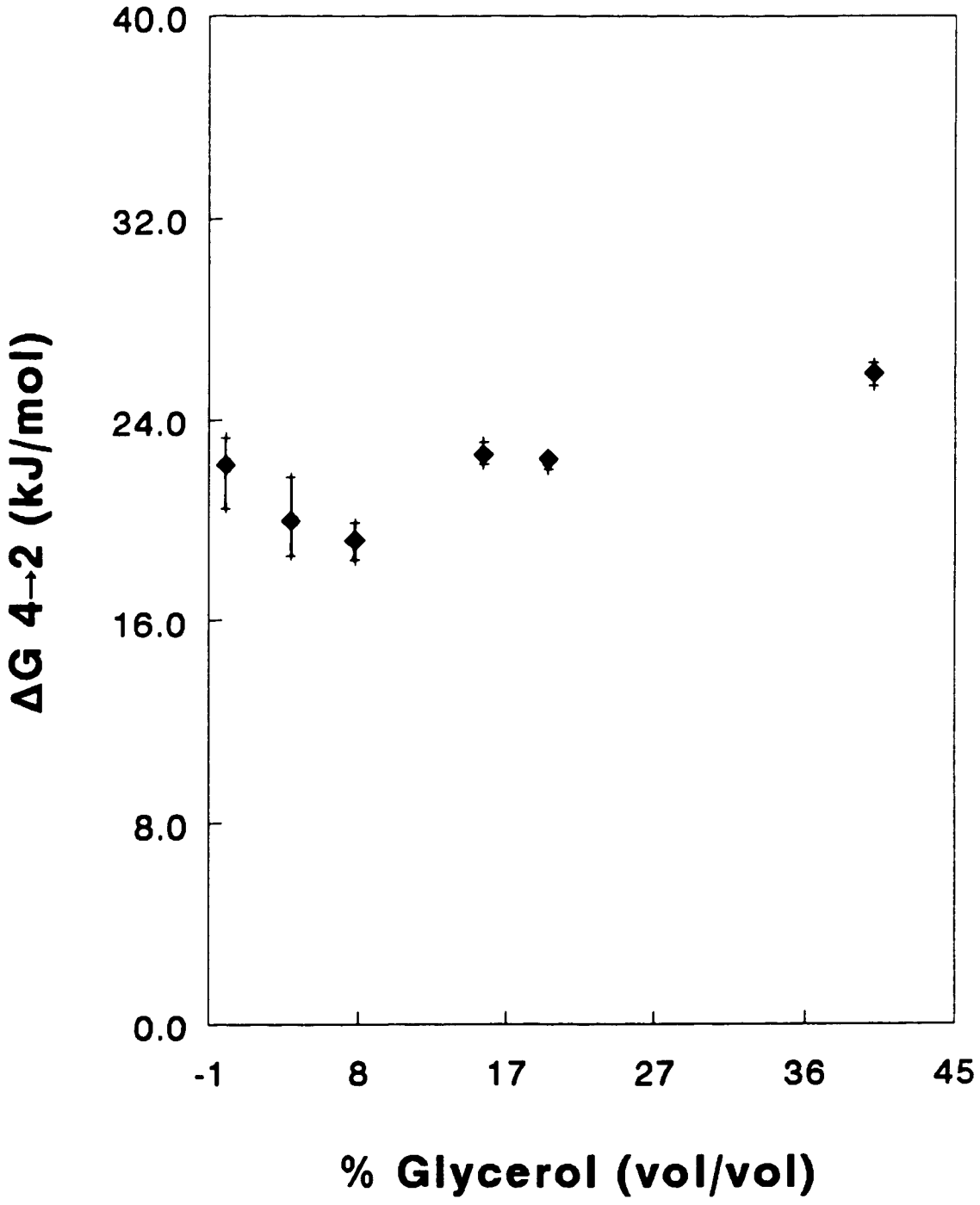
[d]Free energy of dissociation of tetramer to dimer calculated from ln K returned by NONLIN holding θ fixed at the molecular weights listed in note d. Values in parentheses are the 65% confidence interval.

[e]Root mean square of the variance of the fit.

Figure 14. Effect of Glycerol Concentration on the Extent of Association of RSV IN. Free energy of dissociation of IN as a function of glycerol concentration. IN at 1.2 mg/ml was equilibrated by centrifugal gel chromatography into 500 mM NaCl, 20 mM Tris, and 2 mM BME. Glycerol was then added to the buffer for a final concentration (vol/vol) of 8%, 16%, 20% or 40%. Protein loading concentrations were 0.3 - 1.2 mg/ml. K_d values were calculated from LnK returned by NONLIN from simultaneous fit to data acquired at rotor speeds of 20,000, 24,000, 28,000 and 36,000. Free energy values were calculated as described in the Materials and Methods section. Error bars represent the 65% confidence level.

Effect of Glycerol Concentration

Figure 14



difference in the model of self-association returned by NONLIN as indicated in Table 7. While many models which included a monomer in the equilibrium were tested to the data using NONLIN, the best-fit association in all cases was a dimer-tetramer equilibrium. The model of association is probably still a monomer-dimer-tetramer, but the monomer-dimer association is complete at a concentration too low to be detected by currently available ultracentrifuge optics. The data presented in Table 7 show that increasing the glycerol concentration results in the presence of a species which is at least as large as a dimer at a concentration which is high enough to be detected by NONLIN. The data also show that the dimer-tetramer formation becomes more favored as glycerol concentration is increased from 8 to 16% There is no significant difference in the association at 16 and 20%. The dimer-tetramer formation again increases slightly when the glycerol is increased from 20% to 40%.

Chaotropic Series

It was noticed while working with IN in previous experiments that the molecule appears to be insoluble in lower salt solutions at the protein concentrations needed to perform ultracentrifugation studies. It has also been seen in previous experiments that the molecule appeared to be more soluble in KCl than in NaCl and that the monomer-dimer and monomer-tetramer associations were stronger in the NaCl solutions. A chaotropic series experiment was therefore

designed to further study the effect of different buffers on the solubility and association properties of RSV IN. This experiment was performed on the cationic portion of the salt using ammonium chloride, lithium chloride and an equimolar mixture of sodium chloride and potassium chloride as the salt in the buffers. The order of the cations used in this experiment as defined by the Hofmeister series are: NH_4^+, K^+, Na^+, Li^+. Thus, ammonium chloride would be the most disruptive of water structure and decrease the solubility of nonpolar substances more than the other buffers, while lithium chloride would affect these properties the least. Due to the fact that differences have been observed in the free energy of association between KCl and NaCl in earlier experiments, an equimolar mixture of these two salts was also used in this experiment.

IN at a concentration of 1.2 mg/ml was equilibrated in the buffers as previously described. Equilibrium sedimentation studies showed that the protein had remained soluble at this concentration in all conditions studied. The data from these experiments were analyzed by NONLIN and used to determine the stoichiometries and K_d values of IN under increasing chaotropic conditions. The data show that the strength of the monomer-dimer association increased as the cation became less chaotropic. The results of these experiments are given in Table 8.

The result of the equimolar mixture of KCl/NaCl is that

TABLE 8

CHAOTROPIC SERIES OF RSV INTEGRASE

Buffer[a]	$K_{d\ 2\text{-}1}$ μM	$\Delta G_{2\text{-}1}$ kJ/mol	$K_{d\ 4\text{-}1}$ M^3x10^{-14}	$\Delta G_{4\text{-}1}$ kJ/mol	$\Delta G_{4\text{-}2}$ kJ/mol	RMS fringes
NH_4Cl	488.4 (362.6; 679.4)	18.8 (17.9; 19.5)	27.3 (24.2; 30.8)	71.2 (70.9; 71.5)	33.6 (31.9; 35.7)	.011
KCl	270.8 (215.1; 344.2)	20.2 (19.6; 20.8)	76.5 (53.9; 110.7)	68.6 (67.7; 69.5)	28.2 (26.1; 30.3)	.010
KCl/ NaCl[b]	139.9 (106.8; 185.2)	21.8 (21.1; 23.6)	7.2 (6.0; 8.6)	74.4 (74.0; 74.9)	30.8 (26.8; 32.7)	.011
NaCl	157.8 (133.1; 188.9)	21.5 (21.1; 22.0)	54.4 (40.7; 73.5)	69.5 (68.7; 70.2)	26.5 (24.7; 28.0)	.009
LiCl	51.1 (43.0; 60.4)	24.3 (23.9; 24.7)	18.7 (14.0; 25.5)	72.1 (71.3; 72.8)	23.5 (21.9; 25.0)	.011

[a]Concentration of all buffers is 20 mM Tris(pH 8.4), 2 mM BME and 250 mM of the indicated salt. All other constants as defined in Table 2.

[b]Equimolar solution of KCl and NaCl.

dimer formation from monomer is the same as NaCl and slightly higher than KCl. The dimer-tetramer formation of this solution is stronger than either the NaCl or KCl solution.

Self-Association of Structural Mutant Molecules

The mutants provided for study were point mutations S85C, H9N, K164A, W61F, T66A, and D121E. The designation of these mutations refers to the original conserved amino acid (first letter), the position of that amino acid in the IN sequence and the substituted amino acid (last letter). The CO4 deletion mutation was also analyzed. Sedimentation equilibrium analysis was performed on all of these mutants in a buffer containing 500 mM KCl. The results of these experiments are given in Table 9. The comparative activity data discussed below is supplied by the laboratory of Dr. Skalka (76).

T66A. This mutant showed activity that was approximately 25% that of wild type. When tested in the ultracentrifuge it formed soluble aggregates in the buffer used. The data sets were not suitable for NONLIN analysis.

D121E. This mutant showed a markedly reduced activity of 1.7% of wild type. As with T66A, this mutant also formed a soluble aggregate and could not be analyzed.

S85C. This residue has been identified as an invariant residue among retroviral integration proteins. One study concluded that conservative replacements of this residue produced severe reductions in both the processing and joining

TABLE 9

INTEGRASE MUTANTS

Mutant[a]	$K_{d\ 2-1}$ μM	$\Delta G_{\ 2-1}$ kJ/mol	$K_{d\ 4-1}$ M^3x10^{-14}	$\Delta G_{\ 4-1}$ kJ/mol	$\Delta G_{\ 4-2}$ kJ/mol	RMS fringes
S85C[b]	N/A	N/A	N/A	N/A	N/A	.011
H9N	151.6 (122.9; 187.0)	21.6 (21.1; 22.2)	33.4 (29.6; 38.0)	70.7 (70.4; 71.0)	27.5 (26.0; 28.8)	.015
K164A	84.9 (70.0; 102.6)	23.1 (22.6; 23.5)	9.5 (8.3; 10.9)	73.8 (73.4; 74.1)	27.6 (26.4; 28.9)	.008
W61F	1482.1 (846.6; 2620.8)	16.0 (14.6; 17.4)	12.9 (12.1; 13.6)	73.0 (72.9; 73.2)	41.0 (38.1; 44.0)	.010
CO4[c]	N/A	N/A	N/A	N/A	N/A	.010
Wild Type[d]	12.2 (9.9; 15.1)	27.8 (27.3; 28.3)	1.7 (1.4; 2.2)	78.0 (77.4; 78.5)	22.4 (20.8; 23.9)	.012

[a]In 500 mM KCl, 20 mM TRIS, 2 mM BME, pH 8.4. All other constants as defined in Table 2.

[b]This mutant did not self-associate. The monomer molecular weight as returned by NONLIN is 32,650 (31,800;34,300).

[c]This mutant did not self-associate. The monomer molecular weight as returned by NONLIN is 18,350 (17,700;19,000).

[d]The reported values in this data set are from a specimen run in parallel with the mutants.

activities compared to wild type IN (38). Activities of different preparations of this mutant have varied, however. While all data seem to indicate a marked reduction in the activity of S85C when compared to wild type, not all preparations of this mutant show any measurable activity. The mutant which was tested in this experiment at a loading concentration of 0.8 mg/ml showed no activity. The results show that this mutant failed to self-associate.

CO4. This is a mutant in which 118 residues were deleted from the carboxy terminus. This mutant shows no activity when tested for processing and joining activities. The sedimentation equilibrium experiments show that this mutant tested at a loading concentration of 1.5 mg/ml does not self-associate.

H9N. The H9N mutant occurs in the HHCC region of the molecule which has been shown to bind zinc. This region also has been implicated in protein-protein interactions. Activity analysis shows that this mutant has processing and joining activity equal to wild type. Sedimentation equilibrium analysis at a loading concentration of 1.1 mg/ml shows that the monomer-dimer association is significantly weaker and the dimer-tetramer association significantly stronger than that of wild type.

K164A. This mutant retains about 10% of enzymatic activity. This amino acid is a member of the highly conserved catalytic region. The data obtained from sedimentation equilibrium at

a loading concentration of 0.6 mg/ml in this experiment show that the monomer-dimer association is significantly reduced while the dimer-tetramer association is enhanced. While the data obtained from NONLIN in this experiment indicate that the self-association of this mutant is not as strong as wild type, it is the least affected of the mutants studied. It has been shown that hydrophobic and basic amino acids are often involved in direct interactions with DNA and suggested that the reduced activity may be due to impaired substrate interactions (51). The data support the fact that this mutant forms dimer with a decreased ΔG. However, no conclusions can be drawn about the effect of this decrease on activity as a result of reduced DNA binding because the substrate was not present in this experiment.

W61F. The Trp-61 residue is the first conservative residue in the catalytic region. Activity data performed on this mutant have been somewhat inconsistent. Ultracentrifuge data at a loading concentration of 0.7 mg/ml show that the dimerization of this mutant is greatly reduced, while tetramer formation from dimer is very significantly enhanced.

HIV Integration Protein

The similarities of the integration mechanism between RSV and HIV are well documented (1-5,10-13). Studies have shown that HIV IN is both necessary and sufficient for the cleavage and joining of HIV DNA into a target DNA (22,34,36). An experiment was designed to determine if HIV IN self-

associated, and to determine the stoichiometry and association constants if self-association occurred.

HIV IN was equilibrated in buffers as previously described. Analysis by NONLIN of the data in the equilibrium ultracentrifugation experiment showed that the protein was aggregating in all conditions studied. Figure 15 shows the molecular weight values calculated from the sigma returned by the data sets at separate speeds from both the 1.0 and 0.5 M NaCl solutions. The strongly negative slope of the line is an indication of a system that is aggregating rather than achieving equilibrium. Figure 16 is a comparison of the molecular weight calculated for HIV IN and RSV IN in 1.0 M NaCl at the speeds studied. Again, the strongly negative slope of the HIV data indicated aggregation, while a system in equilibrium is indicated by the RSV data whose slope approached zero.

The data show that sedimentation equilibrium techniques cannot be applied to HIV IN at this time due to its impurity. There have recently been two monoclonal columns developed which seem to better purify HIV IN preparations (73). These new preparations may provide material which will be suitable for analysis using these techniques.

Figure 15. Molecular Weight of HIV IN as a Function of Rotor Speed. The solid line represents HIV IN at a loading concentration of 0.4 mg/ml in a buffer of 1 M NaCl, 1 mM EDTA, 20 mM Tris (pH 7.4) and 2 mM BME. The broken represents HIV IN at the same concentration in the same buffer except that the NaCl concentration is 0.5 M. The molecular weight data is calculated from the sigma returned by NONLIN at each speed studied.

HIV Integrase

Figure 15

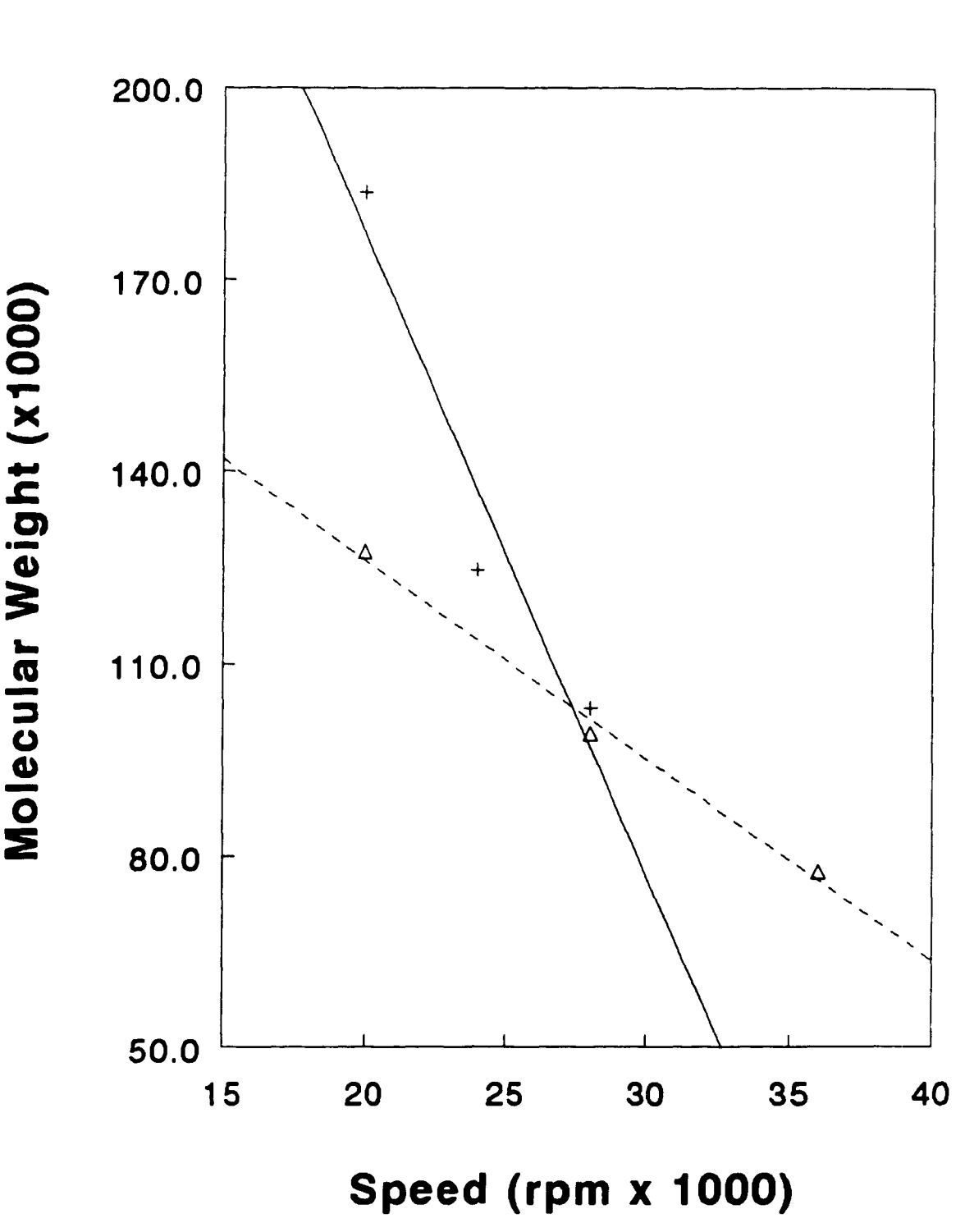
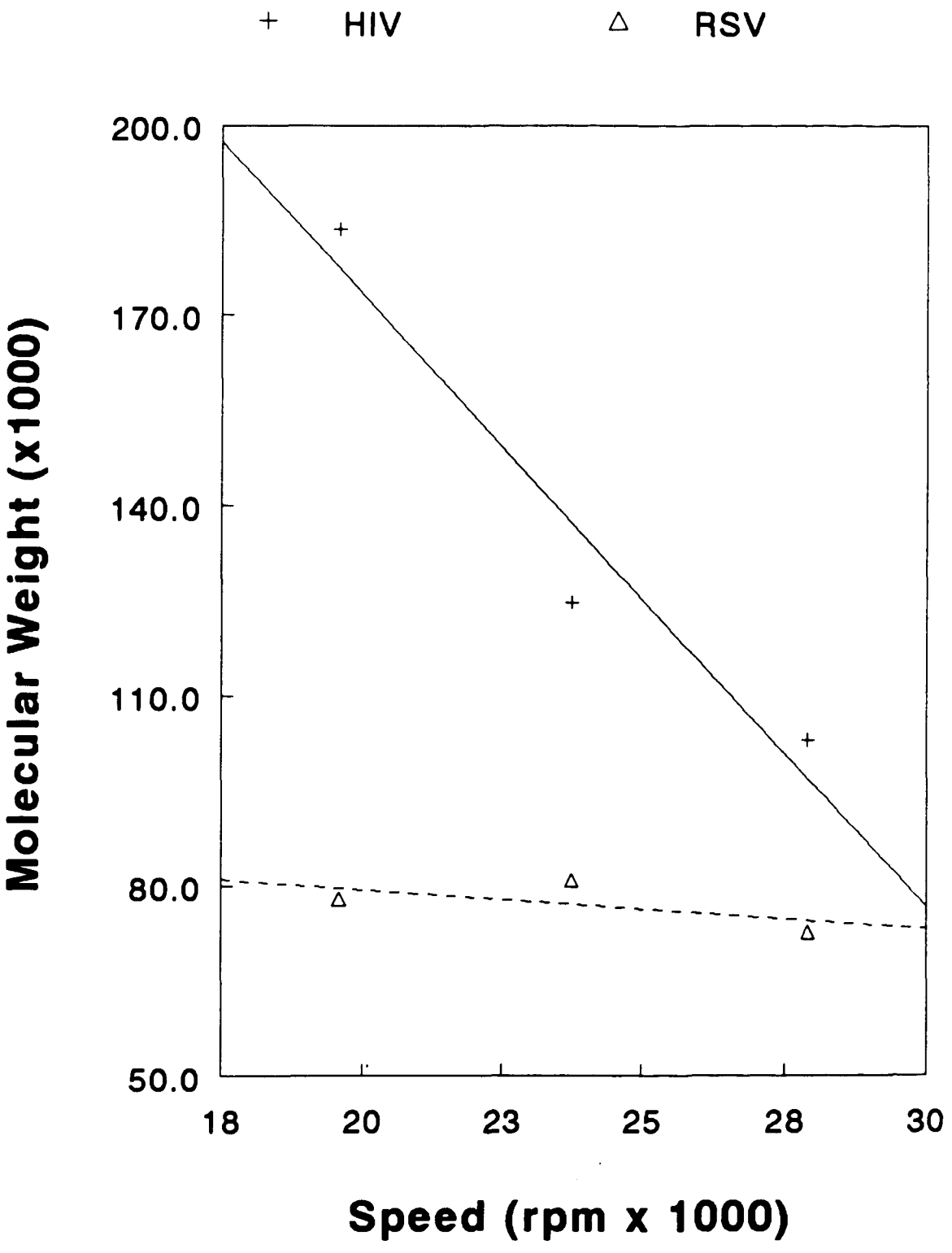


Figure 16. Molecular Weight of RSV and HIV IN as a Function of Rotor Speed. The solid line represents HIV IN at a loading concentration of 0.4 mg/ml in a buffer of 1 M NaCl, 1 mM EDTA, 20 mM Tris (pH 7.4) and 2 mM BME. The broken line represents RSV IN at a loading concentration of 1.5 mg/ml in a buffer of 1 M NaCl, 20 mM Tris (pH 8.4) and 2 mM BME. The molecular weight data is calculated from the sigma returned by NONLIN at each speed studied.

Integrase 1.0M NaCl

Figure 16



DISCUSSION

The results of the experiments show that the best-fit model for wild type RSV IN under all glycerol-free conditions studied is a reversible monomer-dimer-tetramer self-association. Since the individual molecules have a moderate positive charge, there must be forces which overcome this positive charge so that the molecules can associate. What is the nature of these forces?

Forces Affecting IN Self-Association

Decreasing the pH from 8.4 to 7.5 had a minimal effect on the monomer-dimer or monomer-tetramer formation. When the pH was lowered to 6.1, however, there was a dramatic effect on both of these associations: the dimer-tetramer association was greatly weakened as the monomer-dimer association was strengthened. Since decreasing the pH will increase the net positive charge on the molecule, it seems unlikely that the contact region responsible for monomer-dimer self-association is the same as that responsible for dimer-tetramer assembly. The conformation of the residues of the dimer sites is such that the forces which aid dimerization are enhanced by the increased charge. These forces could include dipole-dipole moments, charge-charge interactions or hydrogen bonding. It is not possible from the existing data to eliminate the possibility that the apparent weakening of the dimer-tetramer

association may be an artifact of the NONLIN fitting program, resulting from anticorrelation of lnK_2 and lnK_4 as it adjusts to the strengthened monomer-dimer association.

The protein tends to salt out as the ionic strength of both NaCl and KCl is decreased. Since the molecule carries a positive charge in all solutions tested, the opposite effect would be expected (74). As the ionic strength is lowered, the protein should be less shielded by the salt ions and repel each other because of the positive charge of the protein molecule. This repulsion should increase the solubility of the molecules in solution. This data provides evidence that another force, perhaps dipole-dipole or charge-dipole, is responsible for the insolubility of the protein in solutions of decreasing ionic strength.

The data from the glycerol experiments indicate that the addition of glycerol to buffer strengthens the dimer-tetramer association. In the case of solutions containing at least 8% glycerol (vol/vol), this association is strengthened to the point that the monomer-dimer association is complete at concentrations too low to be detectable by present sedimentation optics. The glycerol effect can perhaps be explained by the nonspecific interaction of glycerol which tends to reduce a thermodynamically unfavorable process by increasing protein self-association and thereby decreasing the surface area of the solvent-protein interaction (75). These results are suggestive of strong hydrophobic interactions

which dramatically increase the monomer-dimer association and significantly increase the dimer-tetramer formation.

The chaotropic experiments show that the monomer-dimer association is weakened by increasingly disruptive cations, and the dimer-tetramer association is strengthened. Increasing the chaotropic ability of the cation disrupts the water structure and decreases the hydrophobic effect in the solution (74). These data support the glycerol results, which indicate that hydrophobic forces are important in the monomer-dimer association. The reason for the strengthening of the dimer-tetramer association as a result of the disruption of the hydrophobic effect may be due to the presence of a separate binding site for dimer-tetramer formation.

The experiments demonstrate that there are significant thermodynamic differences in the self-association of IN in buffers containing NaCl and KCl. IN is also more soluble in KCl than NaCl. These observations may be due to the fact that IN has a specific binding site for one of these ions, or that it has a binding site for which one of these cations has a greater affinity. It cannot be concluded from these experiments which of these ions has the greater binding affinity.

A review of the data presented in the graphs constructed from the ΔG values of the ionic strength experiments (Figures 8,9,11,12) demonstrates that the trend is for the ΔG values for dimer-tetramer dissociation to become larger than the

corresponding ΔG values for the monomer-dimer dissociation as the ionic strength of the buffer is decreased. The dimer-tetramer association becomes stronger that the monomer-dimer association in the lower ionic strength solutions at both pH 8.4 and 6.1. This could have implications for the association properties of IN in physiological conditions.

A comparison of the ΔG values of corresponding NaCl ionic strength (Tables 2 and 4) indicates that as the pH is changed from 6.1 to 8.4 a change in the dissociation energy ($\Delta\Delta G$) can be demonstrated. This change is relatively small, inconsistent and nonsystematic. A similar trend is seen with the NaCl data in Tables 3 and 5.

Thermodynamics of Structural Mutants

Does mutant activity correlate with thermodynamic properties? The mutant forms of IN all exhibited self-association behavior in the ultracentrifuge that differed significantly from that of the wild type protein. It is important to note that the decreased association of any of these mutant proteins could be due to the fact that the mutation has altered the protein structure. Since is has been demonstrated that multimerization is required for activity, any mutation of the protein which reduces its multimerization potential may produce a significant decrease in activity. Thus, a likely possibility is that the resulting decrease in self-association activity is attributed to the disruption of wild type structure created by the mutation of the protein,

not to the specific mutated amino acid or deleted region. The data demonstrate that the activities of these manipulated proteins, even if only a single amino acid substitution, cannot be linked to the association constants. The monomer-dimer associations of each mutant for which data is available is decreased while the dimer-tetramer formation is increased. Since it is known that self-association is required for IN function, any observations concerning domain function must be evaluated in light of the fact that the protein may not be associating for proper activity.

It is unknown why H9N demonstrates activity which is 100% of the wild type protein, yet demonstrates a 10-fold lower self-association. One possible explanation could be that the increased protein and salt concentration required for the thermodynamic experiments affects this mutation more significantly than the other mutations studied, although no data is currently available to support this hypothesis.

Model of IN Association

What is the most plausible model of IN association? These studies show that a monomer-dimer-tetramer self-association exists, but the actual model for association is unknown. Is there a single binding site for association, or are there two separate sites? The molecule also contacts DNA in least four different places (two in the viral DNA and two in the host DNA). How many binding sites does each IN have for this to occur?

The thermodynamic data support a model for two separate protein binding sites: the protein first dimerizes at one site and then the two dimers form a tetramer through contacts at a second distinct site. This can be envisioned through the use of the C2 symmetry model (74). This symmetry is used in the construction of the possible model of IN binding which is presented in Figure 17. IN is a DNA-binding protein, and no model can be accurately predicted without information as to its DNA binding characteristics. Historically, it has been difficult to determine the thermodynamic interactions of protein and DNA due to the spectral overlap of the two with the absorbance optics available on the ultracentrifuge. Recently, a method has been developed which utilizes 5-hydroxytryptophan (5-OH Trp) expressed *in vivo* using an E. coli tryptophan auxotroph (78). The absorbance of 5-OH Trp is far enough removed from that of DNA to allow study of a heteroassociating DNA-protein system. Efforts are currently underway to grow and isolate 5-OH Trp IN. The availability of this enzyme may help to further define the association properties of retroviral IN.

Conclusion

The study of the thermodynamic behavior and association models of RSV IN will help in understanding the mechanism of its function. This knowledge is important because integrases, due to the variety of their biochemical functions, are considered possible targets of therapeutic intervention in the

prevention and treatment of disease. The understanding of these functions is especially important as the search continues for pharmacological agents capable of preventing the infectious process of HIV-1. The results of this study suggest that perhaps this could be accomplished by inhibiting the multimerization of the enzyme.

Figure 17. A Possible Model of IN Self-Association.

A. Monomer IN binds to each end of the reverse transcribed double-stranded DNA.

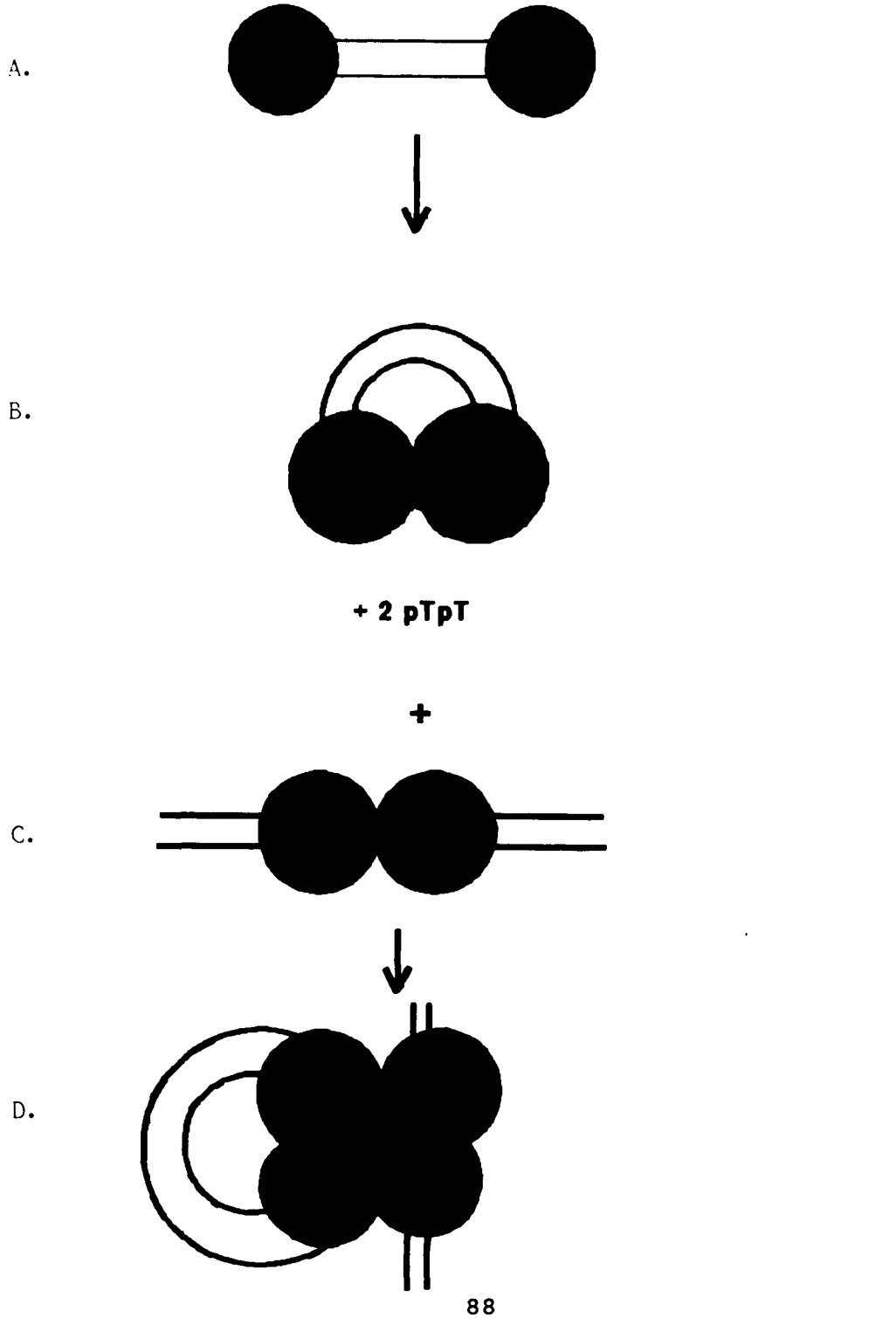
B. Dimerization of IN creates a conformational change in the viral DNA. The removal of the two terminal dinucleotides follows this conformational change.

C. Two separate molecules of IN bind to host DNA forming a dimer.

D. IN tetramer formation results as viral DNA is incorporated into host cell DNA.

MODEL OF IN SELF-ASSOCIATION

Figure 17



APPENDIX A

Figure 18

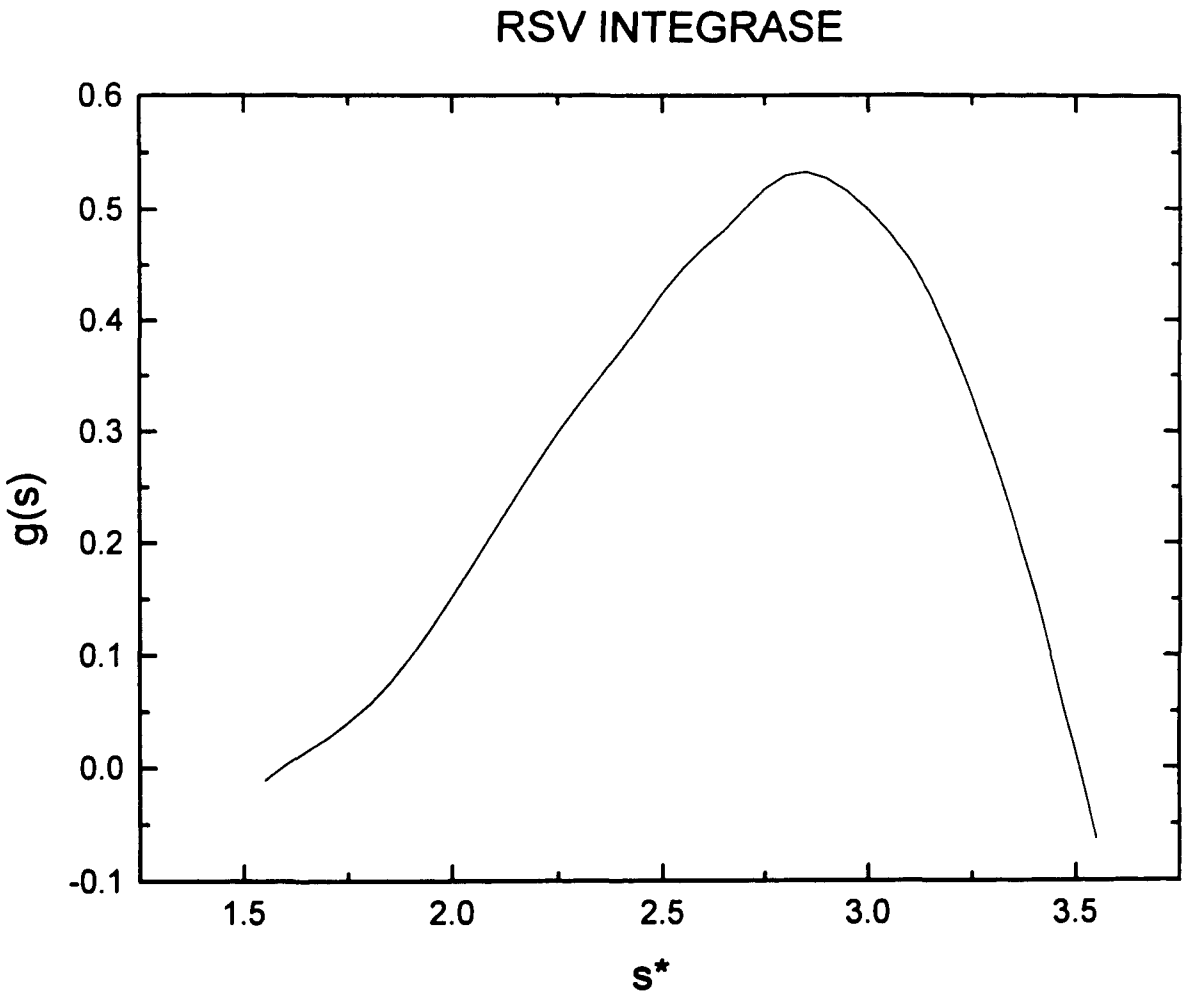


Figure 18. Sedimentation Velocity Experiment. This graph represents the results of a sedimentation velocity experiment performed on RSV IN protein prep 15. 1.2 mg/ml of IN was equilibrated by column gel chromatography into a buffer of 0.5 M NaCl, 20 mM Tris (8.4) and 2 mM BME. The sedimentation velocity run was performed on the Beckman XLA in a double sector cell at 23.3 degrees C. Rotor speed was 60,000 rpm.

The results of this experiment show the high purity of the IN as evidenced by the single peak on the graph. The shoulder on the graph is suggestive of a self-associating species.

APPENDIX B

Procedure for Centrifugal Gel Buffer Exchange for RSV IN.

1. Swell coarse G-25 beads (available from Pharmacia) in water or the buffer to be exchanged for overnight or for 2 hours in a boiling water bath. If water is used to swell the Sephadex, at least five volumes of the exchange buffer should be used to equilibrate the column before use.

2. Disposable centrifugal filtering columns, available from Schleicher and Schuell are used for this procedure. The ones that work best for IN are the 0.2 μm cellulose acetate membrane (Blue top).

3. It is advisable to make several more columns than needed because cracking or unsatisfactory packing of beads during centrifugation occurs occasionally.

4. Remove the sterile sample collector and cover with the cap provided. Replace the collector on the end of the column with an autoclaved eppendorf tube. You may leave the cap attached to the eppendorf tube.

5. Load the columns using a glass pasteur pipet or a disposable plastic bulb syringe. Pack each column with a bead volume of 3 to 5 milliliters; the longer the column, the greater the volume of sample that can be exchanged. Since the columns are to be centrifuged, they need to be poured as equally as possible so they can balance in the centrifuge. Generally, the sample to be desalted can have a volume of 10% of the packed bed volume.

6. The extra buffer/water will drain from the Sephadex through the filter and collect in the eppendorf tube. Empty and replace the eppendorf tube as needed. The use of a vacuum greatly speeds up this process. Put an eppendorf tube with the closed end cut off on the end of the vacuum tube to form a tight seal with the column.

7. Equilibrate the column with buffer through gravity filtration or using the vacuum. When finished with this step the buffer should be just at the top of the beads.

8. Use the Sorvall Centrifuge in with the SS34 rotor. Make sure that the columns are balanced.

9. Replace the eppendorf tubes on the bottom of the centrifuge column with a new sterile labelled eppendorf tube. Load the rotor.

10. The tubes are spun at 2000 rpm for **exactly two minutes.** Start the timer as soon as you push the "on" button - don't wait for the centrifuge to come up to speed.

11. After the centrifuge has stopped, carefully remove and examine the columns. They should appear dry and compressed. The beads will have angled tops and the sides will have pulled away from the walls. Discard the column if there are any horizontal breaks in the beads, but other than that they're ready for use.

12. Remove the eppendorf tube with the buffer in it and cap immediately. This is the buffer to be used in the reference columns when loading centrifuge cells.

13. Place the properly labelled sterile collecting tube that came with the column back on the bottom of the column.

14. Slowly load the protein onto the center of the Sephadex, using a P-200 Pipetperson.

15. Repeat step 10.

16. After the centrifuge has stopped, remove the tubes from the column. Cap with the sterile caps that came with the tubes. The solution in this tube is the protein in the new buffer.

Verification of Centrifugal Gel Buffer Exchange Procedure.

The above procedure was verified by the following steps:

1. The storage buffer for IN is in 40% glycerol. The presence of glycerol will prolong the time required for a specimen to come to equilibrium in the ultracentrifuge. This provided a good marker to test this method while doing a sedimentation experiment. IN was equilibrated using both dialysis and the above method in 0.5 M NaCl, 20 mM Tris (8.4) and 2 mM BME. Both specimens were run in the Beckman Model E. The equilibrium time for both specimens, one and a half hours, was the same. The conclusion could be made that the glycerol is left behind in the buffer exchange.

2. The loading concentrations of both of the above samples was 1.2 mg/ml. The purity of the IN sample had been established previously by sedimentation velocity. The protein concentration of the samples in the ultracentrifuge was determined using the "hinge method" described in the methods

section. The protein concentrations were 1.1 mg/ml in both cells. This showed that the protein was in the sample after the centrifuge gel exchange and that the protein recovery is high for this procedure.

3. NONLIN analysis was performed on both samples in the above experiment. The model of association for both specimens was a monomer-dimer-tetramer with similar K_d values. It has been shown that both glycerol and a change in ionic strength will affect these values. Since the values were similar, it appears that they are in a similar environment.

4. An experiment was performed using the colored solutions that are used to calibrate the pH meters. These solutions were used to represent a storage buffer since they would provide visual evidence of contamination. Columns were prepared as for an IN buffer exchange with 0.5 M NaCl, 20 mM Tris (8.4) and 2 mM BME. The columns were spun to remove the buffer. The colored solutions were then applied to the columns in place of a protein sample. The columns were respun. After the completion of this centrifugation step, the colored buffers had remained in the top portion of the column. Thus it appears that the storage buffer is left behind in this experiment.

APPENDIX C

The procedure which follows on the next seven pages was written in the laboratory of Dr. Anna Marie Skalka by George Merkel of Fox Chase Cancer Research Center, Philadelphia, PA. This is the procedure used to purify the RSV IN used in these experiments. All purified RSV IN was kindly provided for these experiments by George.

Procedure for Purifying RSV IN from E. Coli

1. A known weight of cells are thawed in 50 mM Tris (7.5), 5 M NaCl, 1% Thiodiglycol (TDG), 10 mM EDTA (Lysis buffer) at 0.2 gm or less cells/ml.
a) From weight of cells determine final vol. Thaw and suspend cells in 1/2 this vol. of 50 mM Tris (7.5).
b) Assume weight of cells equals vol. of cells.
c) Add vol. of TDG and 0.5 M EDTA necessary for indicated concentration at final vol.
d) Add NaCl as a solid. At 5M, the NaCl will displace 104 ml/L solution.
e.g. 104 ml/L x 0.025L = 2.62 ml displaced by the solid NaCl in a 25 ml vol.
e) Subtract the vol. of the cells, TDG, 0.5M EDTA and the vol. displaced by the NaCl from the final vol. Then subtract the vol. of buffer already added to the cells to suspend them. The difference is the vol. necessary to achieve the final vol.

2. Stir 30 min./4 deg. C to dissolve NaCl.

3. Lyse cells using a French press, two passes, at 20,000 psi, 4 deg. C. Collect lysate in a tared beaker.
Sonication can be used, but I prefer the press as no heat is generated in the lysate.

4. Weigh lysate. DO NOT centrifuge to clear. Add solid PEG 8K to a final concentration of 6% w/w and solid Dextran T500 to a final concentration of 4% w/w. Mix over night at 4 deg. C.
The Dextran T500 goes into solution with difficulty. This is why the overnight mixing, to ensure it dissolves completely and equilibrium is achieved. The lysate and the polymers are quire viscous, so a good stirrer must be used.
The polymers can be added as stock solutions, 30% w/w PEG 8K, and 20% w/w Dextran T500 in Lysis buffer. If you do it this way with the stock solutions, mixing 1-2 hours is sufficient to achieve equilibrium. I have had no luck in making up these stocks, as one or both of the polymers does not go into solution completely, and I've tried 3-4 times.

5. At this time make up the "wash". To a vol. of lysis buffer equal to 1/2 the weight of the lysate add PEG 8K to 6% w/w, and Dextran T500 to 4% w/w. Mix this overnight at 25 deg. C.

6. After mixing overnight, at 4 deg. C., spin the lysate and the "wash" 30K x g/15 min/4 deg C.

The top (PEG) phase contains most of the proteins in the lysate, including IN, the bottom (Dextran) phase contains most of the DNA. The bacterial debris are found at the interface between the two phases.

Remove the top (PEG) phase from the lysate and save (PEG #1). Then remove the top phase from the "wash" and add it to the bottom phase from the lysate. Mix 1 hour at 4 deg. C to reextract the bottom phase. Discard the "wash" bottom phase.

7. After mixing 1 hour spin the 2nd extraction as in (6). Remove the top (PEG) phase (PEG #2) and combine with PEG #1. Discard the bottom phase.

8. check the conductivity of the combined PEG phases to verify the NaCl concentration.

9. Determine the amount of P-11 phosphocellulose needed as follows:
a) Prepare Whatman P-11 cellulose phosphate ion exchanger as per Green et at (P.J. Green et al, Nuc. Acids Res., 5(7): 2373, 7/78).
b. Equilibrate P-11 with 50 mM Tris (7.5), 0.2M NaCl, 1% TDG, 0.1 mM EDTA, 10% glycerol (0.2 TTEG). If working with CO4 or its derivatives, P-11 should be equilibrated with 0.1 M TTEG.
c. 5 ml packed P-11 is used per gram cells lysed. The settled vol. of P-11 is larger than that obtained when packing in a column, so one must correct for this.
wt. cells x 5 ml packed P-11/gm cells x 1.13 = vol settled P11
e.g. 10 gm x 5 ml packed P-11/gm cells = 50 ml x 1.13 = 56.5 ml settled P-11 needed to process 10 gm cells.
It takes a long time for P-11 to settle completely (24 hours), so spin P-11 slurry for 1 min. at about 200 x g to accelerate the process.

10. Dilute combined PEG phases to 0.2 M NaCl with 0.0 M TTEG. you will increase your vol. by about 25 fold. If working with CO4 or its derivatives, dilute PEG phases to 0.1 M NaCl. Add P-11 to this and stir occasionally for 30 min at 4 deg C to keep the P-11 suspended.

11. Collect P-11, at 4 deg. C., by filtration using a Buchner funnel or a coarse frit glass filter funnel. DON'T LET P-11 DRY OUT!

Wash P-11 in the funnel 3 times with a vol of 0.2 M TTEG equal to that of the packed P-11.

Elute P-11 in the funnel 3 times with a vol. of 1.2 M TTEG equal to that of the packed P-11.

P-11 elutes can be stored at -70 deg. C for 1-2 months.

Run SDS PAGE (12%) to estimate the amount of IN present in the P-11 eluates. My experience is that the first elution fraction will be about 20% IN. Also check the FT and the wash, to make sure you didn't overload the P-11.

12. Purification of IN with MAb column
MAb-A1 is used for WT RSV IN
MAb-B1 is used for CO4 and its derivatives

Determine the amount of MAb linked to the matrix. We use Amino Link coupling gel from Pierce. As the molar ratio of Mab:IN is 150 Kd:32Kd = 4.7, one pass over the column will give you only an amount of IN equal to 1/5 the mass of Mab linked to the column. Multiple passes over the column will be necessary to purify all the IN. The Mab col. is run at room temperature.

13. Thaw P-11 eluate, take an aliquot containing an amount of IN equal to the Mab col. capacity.

Check its conductivity, and dilute, if necessary, to 0.5 M NaCl with either 0.0 M TTEG or 1X PBS.

14. a) Apply eluate to Mab col. at 30 ml/hr/cm*2 (matrix is Sepharose 6B, don't crush it). Collect flow through and save at 4 deg C.
b) Wash col. with 10 col. vol. 25 mM NaPO4 (7.4), 0.5M NaCl. collect wash and save at 4 deg. C.
c) Elute col. with 0.1 M glycine (3.0), 0.5 M NaCl. Most of the protein will elute over the first 4 col. vol., with the peak at the second col. vol. The eluate is neutralized by collecting into an equal vol. of cold 1 M Tris (7.5), 0.5 M NaCl. Do this as follows:

Place a col. vol. of cold 1 M Tris (7.5), 0.5 M NaCl in a beaker in an ice bath with stirring. Elute col. with 4 col. vol. 0.1 M glycine (3.0), 0.5 M NaCl. Collect eluate into the beaker containing the cold neutralizing buffer, with stirring. This way the eluate is neutralized immediately, and diluted so no precipitation can occur.

Repeat as above with another 6 col. vol. of elution and neutralization buffer. This second elution will contain virtually no protein, but insures that the col is cleaned off.
d) Wash col. with 10 col. vol. 25 mM NaPO4 (7.4), 0.5 M NaCl, if you are going to use the col. the next day. If not, wash the col. with 10 col. vol. 1X PBS, 0.02% Na azide, and store at 4 deg. C.

Flow through and wash should be checked on a SDS gel for the presence of IN. If a significant amount of IN is found in the FT and wash, simply pass them over the col. again.

15. Take A280 of elution fractions.
For WT RSV IN, 1A280u/ml = 0.4 mg/ml
For CO4 and derivatives, 1A280 u/ml = 0.6 mg/ml
These are approximate only!
Calculate the amount of IN present in the elution fractions.

16. The solubility for WT RSV IN is 1.87 mg/ml in 0.5 M NaCl storage buffer, or 4.5 mg/ml in 1.0 M NaCl storage buffer.

For CO4 and derivatives, the upper solubility limit has not been established. I have concentrated them to 3.0 mg/ml in 0.5 M NaCl storage buffer with no precipitation.

Protein concentrations above are uncorrected BioRad protein estimations.

For WT RSV IN we have a correction factor, BioRad estimation x 0.8 = true protein concentration. This should only be used on final pools.

17. Now you know the amount of IN present in the MAb col. eluate. You wish to concentrate it to its highest useful concentrations, so...

Amount IN/max. final conc. = final vol x 3 = intermediate vol.

Using an Amicon stirred cells with a YM 20 membrane, concentrate eluate to the calculated intermediate vol.

Place this vol. in dialysis vs 3 (1L) changes of 50 mM HEPES (8.1), 0.5/1.0 M NaCl, 1% TDG, 0/1 mM EDTA, 40% glycerol (Storage buffer).

During dialysis the vol. will decrease by a factor of three.

18. Store final pool at -20 deg C. You will have to run the Mab col. a number of time to purify all the P-11 eluate. At -20 deg. C the pool will not freeze, so you can if you want, pool them all together at the end.

19. IN keeps for at least 3 yrs. at -70 deg. C, and can be stored in liquid nitrogen. At -20 deg. C., IN keeps at least 6 months.

IN is susceptible to freeze/thaw, after about 6 cycles it dies. This problem can be avoided by keeping working vols. at -20 deg. C, and thawing fresh tubes as needed. This way the protein is free/thawed only once or twice.

p32(19) Final Summary

CUT	ug/ml	Vol,ml	Total mg	Corr ug/ml	Corr Total mg	Vo/pMol IN	M NaCl	Corr pMol/ul
P	1406.3	39.54	55.6	1125.04	44.48	0.16446	0.5	35.16
S	1424.7	36.64	52.2	1139.73	41.76	0.15438	0.5	35.17
T	4004.6	11.65	46.65	3203.72	37.32	0.15197	1	100.12
J	1876.7	6.8	12.76	1501.36	10.21	ND	0.5	46.9
K	1787.9	7.3	13.05	1430.32	10.44	ND	0.5	44.7
		Sum P,S,&T=	154.45mg		123.56mg			
		Sum ALL=	180.26mg		144.21mg			

REFERENCES

1. Brown, P.O. (1990) Integration of Retroviral DNA. In: *Current Topics in Microbiology and Immunology*. Springer-Verlag, Berlin-Heidelberg, 157:19-48.

2. Grandgenett, D.P., Mumm, S.R. (1990) Unraveling Retrovirus Integration. *Cell*. 60:3-4.

3. Kulkosky, J., Skalka, A.M. (1990) HIV DNA Integration: Observations and Inferences. *Journal of AIDS*. 3:839-851.

4. Kulkosky, J., Skalka, A.M. (1993) Molecular Mechanism of Retroviral DNA Integration. *Pharmacology and Therapeutics*. in press.

5. Varmus, H., Brown, P. (1989) Retroviruses. In: Berg, D., Howe, M. eds. *Mobile DNA*. Washington,DC: American Socity for Microbiology, 53-108.

6. Leis, J., Baltimore, D., Bishop, J.M., Coffin, J., Fleissner, E., Goff, S.P., Oroszlan, S., Robinson, H., Skalka, A.M., Temin, H.M., Vogt, V. (1988) Standardized and Simplified Nomenclature for Proteins Common to All Retroviruses. *Journal of Virology*. 62:1808-1809.

7. Cooper, G. (1990). *Oncogenes*. Jones and Bartlett, Boston, 35-49.

8. Kunj, H.J., Boerkoel, C., Carter, T.H. (1991) Retroviral Mutagenesis of Cellular Oncogenes: A Review with Insights into the Mechanisms of Insertional Activation. In: Kung, H.J., Vogt, P.K. eds. *Current Topics in Microbiology and Immunology*. Springer-Verlag, Berlin-Heidelberg, 171:1-25.

9. Grandgenett, D.P., Ajaykumar, C.V., Schiff, R.D. (1978) A 32,000-Dalton Nucleic Acid-binding Protein from Avian Retravirus Cores Possesses DNA Endonuclease Activity. *Virology*. 89:119-132.

10. Katz, R.A., Merkel, G., Kulkosky, J., Leis, J., Skalka, A.M. (1990) The Avian Retroviral IN Protein Is Both Necessary and Sufficient for Integrative Recombination In Vitro. *Cell*. 63:87-95.

11. Roth, M.J., Tanese, N., Goff, S.P. (1988) Gene Product of Moloney Murine Leukemia Virus Required for Proviral Integration is a DNA-binding Protein. *Journal of Molecular Biology*. 203:131-139.

12. Mirsa, T.K., Grandgenett, D.P., Parsons, J.T. (1982) Avian Retrovirus pp32 DNA-binding Protein 1. Recognition of Specific Sequences on Retrovirus DNA Terminal Repeats. *J. Virol.* 44:330-343.

13. Pryciak, P.M., Varmus, H.E. (1992) Nucleosomes, DNA-Binding Proteins, and DNA Sequence Modulate Retroviral Integration Site Selection. *Cell.* 69:769-780.

14. Bowerman, B., Brown, P.O., Bishop, J.M., Varmus, H. A. (1989) Nucleoprotein Complex Mediates the Integration of Retroviral DNA. *Genes and Development.* 3:469-478.

15. Goff, S.R. (1992) Genetics of Retroviral Integration. *Annu. Rev. Genet.* 26:527-44.

16. Majors, J. (1990) The Structure and Function of Retroviral Long Terminal Repeats. *Current Topics in Microbiology and Immunology.* Springer-Verlag, Berlin-Heidelberg. 157:49-92.

17. Cobrinik, D., Katz, R., Terry, R., Skalka, A.M., Leis, J. (1987) Avian Sarcoma and Leukosis Virus *pol*-Endonuclease Recognition of the Tandem Long Terminal Repeat Junction: Minimum Site Required for Cleavage Is Also Required for Viral Growth. *Journal of Virology.* 61:1999-2008.

18. Colicelli, J., Goff, S.P. (1988) Sequence and Spacing Requirements of a Retrovirus Integration Site. *J. Mol. Biol.* 199:47-59.

19. Katzman, M., Katz, R.A., Skalka, A.M., Leis, J. (1989) The Avian Retroviral Integration Protein Cleaves the Terminal Sequences of Linear Viral DNA at the In Vivo Sites of Integration. *Journal of Virology.* 63:5319-5327.

20. Panganiban, A.T., Temin, H.M. (1983) The Terminal Nucleotides of Retrovirus DNA are Required for Integration but not Virus Production. *Nature.* 306:155-160.

21. Roth, M.J., Schwartzberg, P.L., Goff, S.P. (1989) Structure of the Termini of DNA Intermediates in the Integration of Retroviral DNA: Dependence on IN Function and Terminal DNA Sequence. *Cell.* 58:47-54.

22. Bushman, F.D., Fujiward, T., Craigie, R. (1990) Retroviral DNA Integration Directed by HIV Integration Protein In Vitro. *Science.* 249:1555-1558.

23. Donehower, L.A., Varmus, H.E. (1984) A Mutant Murine Leukemia Virus with a Single Missense Codon in *pol* is Defective in Function Affection Integration. *Proc. Natl. Acad. Sci. USA.* 81:6461-6465.

24. Hippenmeyer, P.J., Grandgenett, D.P. (1984) Requirement of the Avian Retrovirus pp32 DNA Binding Protein Domain for Replication. *Virology*. 137:358-370.

25. Johnson, M.S., McClure, M.A., Feng, D.F., Gray, J., Doolittle, R.F. (1986) Computer Analysis of Retroviral *pol* Genes: Assignment of Enzymatic Functions to Specific Sequences and Homologies with Nonviral Enzymes. *Proc. Natl. Acad. Sci. USA*. 83:7648-7652.

26. Quinn, T.P., Grandgenett, D.P. (1989) Avian Retrovirus Integration Protein: Structure-Functional Analysis of Viable Mutants. *Virology*. 173:478-488.

27. Khan, E., Mack, J.P.G., Katz, R.A., Kulkosky, J., Skalka, A.M. (1991) Retroviral Integrase Domains: DNA Binding and the Recognition of LTR Sequences. *Nucleic Acids Research*. 19:851-860.

28. Brown, P.O., Bowerman, B., Varums, H.E., Bishop, J.M. (1989) Retroviral Integration: Structure of the Initial Covalent Product and its Precursor, and a Role for the Viral IN Protein. *Proc. Natl. Acad. Sci. USA*. 86:2525-2529.

29. Ju, G., Skalka, A.M. (1980) Nucleotide Sequence Analysis of the Long Terminal Repeat (LTR) of Avian Retroviruses: Structural Similarities with Transposable Elements. *Cell*. 223:379-386.

30. Pauza, C.D. (1990) Two Bases Are Deleted from the Termini of HIV-1 Linear DNA During Integrative Recombination. *Virology*. 179:886-889.

31. Terry, R., Soltis, D.A., Katzman, M., Cobrinik, D., Leis, J., Skalka, A.M. (1988). Properties of Avian Sarcoma-Leukosis Virus pp32-Related *pol*-Endonucleases Produced in *Escherichia coli*. *Journal of Virology*. 62:2358-2365.

32. LaFemina, R.L., Callahan, P.L., Cordingley, M.G. (1991) Substrate Specificity of Recombinant Human Immunodeficiency Virus Integrase Protein. *Journal of Virology*. 65:5624-5630.

33. Kulkosky, J., Katz, R.A., Skalka, A.M. (1990) Terminal Nucleotides of the Preintegrative Linear Form of HIV-1 DNA Deduced from the Sequence of Circular DNA Junctions. *Journal of AIDS*. 1990:852-858.

34. Bushman, F.D., Craigie, R. (1991) Activities of Human Immunodeficiency Virus (HIV) Integration Protein *in vitro:* Specific Cleavage and Integration of HIV DNA. *Proc. Natl. Acad. Sci. USA*. 88:1339-1343.

35. Leavitt, A.D., Rose, R.B., Varmus, H.E. (1992) Both Substrate and Target Oligonucleotide Sequences Affect In Vitro Mediation by Human Immunodeficiency Virus Type 1 Integrase Protein Produced by *Saccharomyces cerevisiae*. *Journal of Virology*. 66:2359-2368.

36. Jones, K.S., Kulkosky, J., Skalka, A.M. (1991) Analyses of HIV Integration Components. In: *Advances in Molecular Biology and Targeted Treatment for AIDS*. Kuma, A., ed. Plenum Press, New York, 21-26.

37. Morris-Vasios, C., Kochan, J.P., Skalka, A.M. (1988) Avian Sarcoma-Leukosis Virus *pol*-endo Proteins Expressed Independently in Mammalian Cells Accumulate in the Nucleus byt Can Be Directed to Other Cellular Compartments. *Journal of Virology*. 62:349-353.

38. Chow, S.A., Vincent, K.A., Ellison, V., Brown, P.O. (1992) Reversal of Integration and DNA Splicing Mediated by Integrase of Human Immunodeficiency Virus. *Science*. 255:723-726.

39. Engelman, A., Bushman, F.D., Craigie, R. (1993) Identification of Discrete Functional Domains of HIV-1 Integrase and their Organization within an Active Multimeric Complex. *The EMBO Journal*. 8:3269-3275.

40. Shih, C.C., Stoye, J.P., Coffin, J.M. (1988) Highly Preferred Targets for Retrovirus Integration. *Cell*. 53:531-537.

41. Mooslehner, K., Karls, U., Harbers, K. (1990) Retroviral Integration Sites in Transgenic Mov Mice Frequently Map in the Vicinity of Transcribed DNA Regions. *Journal of Virology*. 64:3056-3058.

42. Scherdin,U., Rhodes, K., Breindl, M. (1990) Transcriptionally Active Genome Regions Are Preferred Targets for Retrovirus Integration. *Journal of Virology*. 64:907-912.

43. Felsenfeld, G. Chromatin as an Essential Part of the Transcriptional Mechanism. *Nature*. 355:219-224.

44. Kornberg, R.D., Lorch, Y. Irrestible Force Meets Immovable Object: Transcription and the Nucleosome. *Cell*. 67:833-836.

45. Stevenson, M., Haggerty, S., Lamonica, C.A., Meier, C.M., Welch, S., Wasiak, A.J. (1990) Integration Is Not Necessary for Expression of Human Immunodeficiency Virus Type 1 Protein Products. *Journal of Virology*. 64:2421-2425.

46. Panganiban, A.T., Temin, H.W. (1984) Circles with Two Tandem LTRs Are Precursors to Integrated Retrovirus DNA. *Cell*.

36:673-679.

47. Brown, P.O., Bowerman, B., Varmus, H.E., Bishop, J.M. (1987) Correct Integration of Retroviral DNA In Vitro. *Cell*. 49:347-356.

48. Fujiwara, T., Mizuuchi, K. (1988) Retroviral DNA Integration: Structure of an Integration Intermediate. *Cell*. 54:497-504.

49. Lobel, L.L., Murphy, J.E., Goff, S.P. (1989) The Palindromic LTR-LTR Junction of Moloney Murine Leukemia Virus Is Not an Efficient Substrate for Proviral Integration. *Journal of Virology*. 63:2629-2637.

50. Vink, C., Yeheskiely, E., van der Marel, G.A., van Boom, J., Plasterk, R.H.A. (1992) Site-specific Hydrolysis and Alcoholysis of Human Immunodeficiency Virus DNA Termini Mediated by the Viral Integrase Protein. *Nucleic Acids Research*. 19:6691-6698.

51. Katz, R.A., Mack, J.P.G., Merkel, G., Kulkosky, J., Zheng, G., Leis, J., Skalka, A.M. (1992) Requirement for a Conserved Serine in both Processing and Joining Activities of Retroviral Integrase. *Proc. Natl. Acad. Sci. USA*. 89:6741-6745.

52. Engelman, A., Mizuuchi, K., Craigie, R. (1991) HIV-1 DNA Integration: Mechanism of Viral DNA Cleavage and DNA Strand Transfer. *Cell*. 67:1211-1221.

53. Katzman, M., Mack, J.P.G., Skalka, A.M., Leis, J. (1991) A Covalent Complex Between Retroviral Integrase and Nicked Substrate DNA. *Proc. Natl. Acad. Sci. USA*. 88:4695-4699.

54. Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P.G., Skalka, A.M. (1992) Residues Critical for Retroviral Integrative Recombination in a Region That Is Highly Conserved among Retroviral/Retrotransposon Integrases and Bacterial Insertion Sequence Transposases. (1992) *Molecular and Cellular Biology*. 12:2331-2338.

55. Bushman, F.D., Engelman, A., Palmer, I., Wingfield, P., Craigie, R. (1993) Domains of the Integrase Protein of Human Immunodeficiency Virus Type 1 Responsible for Polynucleotidyl Transfer and Zinc Binding. *Proc. Natl. Acad. Sci. USA*. 90:3428-3432.

56. Leavitt, A.D., Shiue, L., Varmus, H.E. (1993). Site-directed Mutagenesis of HIV-1 Integrase Demonstrates Differential Effects on Integrase Functions *in Vitro*. *The Journal of Biological Chemistry*. 268:2113-2119.

57. Vink, C., Oude Groeneger, A.A.M., Plasterk, R.H.A. (1993) Identification of the Catalytic and DNA-binding Region of the Human Immunodeficiency Virus Type 1 Integrase Protein. *Nucleic Acids Research*. 21:1419-1425.

58. Van Gent, D.C., Vink, C., Oude Groeneger, A.A.M., Plasterk, R.H.A. (1993) Complementation Between HIV Integrase Proteins Mutated in Different Comains. *The EMBO Journal*. 12:3261-3267.

59. Murphy, J.E., Goff, S.P. (1992) A Mutation at One End of Moloney Murine Leukemia Virus DNA Blocks Clevage of Both Ends by the Viral Integrase In Vivo. *Journal of Virology*. 66:5092-5095.

60. Johnson, M.L., Correia, J.J., Yphantis, D.A., Halvorson, H.R. (1981) Analysis of Data from the Analytical Ultracentrifuge by Nonlinear Least-Squares Techniques. *Biophysical Journal*. 36:575-588.

61. Christopherson, R.I., Jones, M.E., Finch, L.R. (1979) A Simple Centrifuge Column for Desalting Protein Solutions. *Anal. Biochem*. 100:184-187.

62. Pharmacia. (1985) *Gel Filtration and Practice*. Rahms i Lund, Sweden.

63. Yphantis, D.A., Waugh, D.F. (1956) Ultracentrifugal Characterization by Direct Measurement of Activity. *J. Phys. Chem*. 60:623-629.

64. Laue, T.M. (1981) Rapid Precision Interferometry for the Analytical Ultracentrifuge. Ph.D. Dissertation. University of Conn., Storrs, CT.

65. Wetlaufer, D.B. (1962) *Adv. Prot. Chem*. 17:303-390.

66. Weast, R.C., ed. (1986) *CRC Handbook of Chemistry and Physics*. 67th ed. CRC Press, Inc. Boca Raton, FL.

67. Laue, T.M., Shah, B.D., Ridgeway, T.M., Pelletier, S.L. (1992) Computer-Aided Interpretation of Analytical Sedimentation Data For Proteins. In: *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Harding, S., Rowe, A., eds. Royal Society of Chemistry, London, 90-125.

68. Luckow, E.A., Lyons, D.A., Ridgeway, T.M., Esmon, C.T., Laue, T.M. (1989) Interaction of Clotting Factor V Heavy Chain with Prothrombin and Prethrombin 1 and Role of Activated Protein C in Regulating This Interaction: Analysis by Analytical Ultracentrifugation. *Biochemistry*. 28:2348-2354.

69. Roark, D.E., Yphantis, D.A. (1969) Studies of Self-Associating Systems by Equilibrium Ultracentrifugation. *Annals New York Academy of Sciences*. 164:245-278.

70. Jones, K.S., Coleman, J., Merkel, G.W., Laue, T.M., Skalka, A.M. (1992) Retroviral Integrase Functions as a Multimer and Can Turn Over Catalytically. *Journal of Biological Chemistry*. 267:16037-16040.

71. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D. (1983) *Molecular Biology of the Cell*. Garland Publishing, New York-London, 286.

72. Personal Communication with George Merkel, Fox-Chase Cancer Center.

73. Personal Communication with Joe Kulkosky, Fox-Chase Cancer Center.

74. Mathews C.K., van Holde, K.E. (1990) *Biochemistry*. Benjamin/Cummings Publishing Co., Redwood City, CA, 50-53.

75. Gekko, K., Timasheff, S.N. (1981) Mechanism of Protein Stabilization by Glycerol: Preferential Hydration in Glycerol-Water Mixtures. *Biochemistry*. 20:4667-4676.

76. Personal Communication with Barbara Mueller, Fox-Chase Cancer Center.

77. Yphantis, D.A. (1960) Rapid Determination of Molecular Weights of Peptides and Proteins. *Annu. N.Y. Acad. Sci.* 88:586-601.

78. Laue, T.M., Senear, D.F., Eaton S., Ross, J.B.A. (1993) 5-Hydroxytryptophan as a New Intrinsic Probe for Investigating Protein-DNA Interactions by Analytical Ultracentrifugation. Study of the Effect of DNA on Self-Assembly of the Bacteriophage λ *c*I Repressor. *Biochemistry*. 32:2469-2472.